

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2001, 19:12:46 ; Search time 13.99 seconds
(without alignments)
532.422 Million cell updates/sec

Title: US-09-170-042A-2
Perfect score: 1756
Sequence: 1 MENPSPAALGKALCALLLA.....NGSPCELEEEAEVPCVNCV 331

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

11 number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1756	100.0	331	2	US-08-799-173A-2
2	1548.5	88.2	330	4	US-09-371-696-2
3	464.5	26.5	802	1	US-07-862-021B-12
4	464.5	26.5	802	1	US-08-313-288B-12
5	464.5	26.5	802	5	PCT-US93-03164-12
6	462.5	26.3	392	2	US-08-799-173A-7
7	462.5	26.3	807	1	US-07-862-021B-10
8	462.5	26.3	807	1	US-08-313-288B-10
9	462.5	26.3	807	5	PCT-US93-03164-10
10	420.5	23.9	132	4	US-09-022-238-2
11	331	18.8	568	1	US-07-862-021B-14
12	331	18.8	568	5	PCT-US93-03164-14
13	309	17.6	53	2	US-08-799-173A-18
14	189	10.8	37	4	US-09-022-238-3
15	181	10.3	37	4	US-09-371-696-3
16	130.5	7.4	56	1	US-07-862-021B-19
17	130.5	7.4	56	5	PCT-US93-03164-19
18	128.5	7.3	50	2	US-08-799-173A-14
19	121.5	6.9	52	2	US-08-799-173A-12
20	103	5.9	1251	5	PCT-US95-02251-3
21	103	5.9	1252	1	US-08-199-780-3
22	103	5.9	1252	2	US-08-316-650-3
23	103	5.9	1253	3	US-08-479-722B-4
24	99.5	5.7	812	1	US-08-446-794A-4
25	96.5	5.5	812	1	US-08-446-794A-2
26	96.5	5.5	812	1	US-08-750-007-3
27	96.5	5.5	812	2	US-08-945-024-2

28 92.5 5.3 912 5 PCT-US95-03747-2 Sequence 2, Appli
29 91.5 5.2 52 2 US-08-799-173A-10 Sequence 10, Appl
30 91.5 5.2 56 1 US-07-862-021B-17 Sequence 17, Appl
31 91.5 5.2 56 5 PCT-US93-03164-17 Sequence 17, Appl
32 90 5.1 53 2 US-08-799-173A-9 Sequence 9, Appl
33 89.5 5.1 56 1 US-07-862-021B-16 Sequence 16, Appl
34 89.5 5.1 56 5 PCT-US93-03164-16 Sequence 16, Appl
35 89.5 5.1 1156 3 US-08-996-083-1 Sequence 1, Appl
36 89.5 5.1 1156 4 US-09-429-516-1 Sequence 1, Appl
37 89.5 5.1 1156 4 US-09-429-516-3 Sequence 3, Appl
38 89 5.1 1257 4 US-09-220-641-3 Sequence 3, Appl
39 87.5 5.0 1184 2 US-08-918-914-1 Sequence 1, Appl
40 87.5 5.0 1184 3 US-08-996-083-3 Sequence 3, Appl
41 86.5 4.9 224 4 US-08-871-572B-11 Sequence 11, Appl
42 85.5 4.9 692 4 US-09-003-574-31 Sequence 31, Appl
43 85.5 4.9 692 4 US-09-003-570-31 Sequence 4, Appl
44 85.5 4.9 695 2 US-08-701-240-4 Sequence 4, Appl
45 85.5 4.9 695 4 US-09-138-236-4 Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-799-173A-2
; Sequence 2, Application US/08799173A
; Patent No. 5871969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, GREGG,
; APPLICANT: PATRICK J. DILLON
; TITLE OF INVENTION: HUMAN NEURONAL ATTACHMENT FACTOR-1
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/799,173A
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDERS A.
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; #)LEUCLE TYPE: protein
; US-08-799-173A-2

Query Match 100.0%; Score 1756; DB 2; Length 331;
Best Local Similarity 100.0%; Pred. No. 3e-166;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MENPSPAALGKALCALLLAATGAGPLGSGESCSARALAKYSITFTCKWSOTAFPKQY 60
Db 1 MENPSPAALGKALCALLLAATGAGPLGSGESCSARALAKYSITFTCKWSOTAFPKQY 60
QY 61 PLFRPPAQWSSLLGAHSSDYSMMWRKNQYVNSGLRDFAEERGEAWALMKETEAAAGALQSV 120

Db 61 PLFRPPAQSLLGAHSDYSWMRKNOYVSNGLRDFAEERGEAWALMKETEAGEALQSV 120
 QY 121 HAVFSAPAVPSTGOTSAAELEVORRHSLVSFVVRIVPSDFVGVDSLDLDCDGRWRREQA 180
 Db 121 HAVFSAPAVPSTGOTSAAELEVORRHSLVSFVVRIVPSDFVGVDSLDLDCDGRWRREQA 180
 QY 181 ALDLYPDAGTDSGTFSSPNFATIPQDVTTEITSSPSHPANFSFYPRLKALPPIARVT 240
 Db 181 ALDLYPDAGTDSGTFSSPNFATIPQDVTTEITSSPSHPANFSFYPRLKALPPIARVT 240
 QY 241 LVRLRQSPRAFPAPVLPVPSRDNEIVDSASVPETPLDCEVSLWSWGLCGHCGRLGTS 300
 Db 241 LVRLRQSPRAFPAPVLPVPSRDNEIVDSASVPETPLDCEVSLWSWGLCGHCGRLGTS 300
 QY 301 RTRVYRVOPANNGPCPELEEEACVDPNCV 331
 Db 301 RTRVYRVOPANNGPCPELEEEACVDPNCV 331

ULT 2

Sequence 2, Application US/09371696
 Patent No. 6287777
 GENERAL INFORMATION:
 APPLICANT: Sytkowski, Arthur J.
 APPLICANT: Yang, Meiheng
 TITLE OF INVENTION: NOVEL NPC-1 GENE THAT IS DIFFERENTIALLY EXPRESSED IN PROSTATE
 TITLE OF INVENTION: TUMORS
 FILE REFERENCE: 01948/053002
 CURRENT APPLICATION NUMBER: US/09/371.696
 CURRENT FILING DATE: 1999-08-10
 EARLIER APPLICATION NUMBER: US 09/022.238
 EARLIER FILING DATE: 1998-02-11
 EARLIER APPLICATION NUMBER: US 08/644.326
 EARLIER FILING DATE: 1996-05-10
 NUMBER OF SEQ ID NOS: 5
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 2
 LENGTH: 330
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-371-696-2

Query Match 88.2%; Score 1548.5; DB 4; Length 330;
 Best Local Similarity 87.6%; Pred. No. 1.2e-145;
 Matches 297; Conservative 6; Mismatches 19; Indels 17; Gaps 2;
 1 MENPSPAALGKALCALLLATLGAAGQPLGGESICSRAPAKYSITFTGKWSQTAFFPKQY 60
 Db 1 MENPSPAALGKALCALLLATLGAAGQPLGGESICSRAPAKYSITFTGKWSQTAFFPKQY 60
 QY 61 PLFRPPAQSLLGAHSDYSWMRKNOYVSNGLRDFAEERGEAWALMKETEAGEALQSV 120
 Db 61 PLFRPPAQSLLGAHSDYSWMRKNOYVSNGLRDFAEERGEAWALMKETEAGEALQSV 120
 QY 121 HAVFSAPAVPSTGOTSAAELEVORRHSLVSFVVRIVPSDFVGVDSLDLDCDGRWRREQA 180
 Db 121 HAVFSAPAVPSTGOTSAAELEVORRHSLVSFVVRIVPSDFVGVDSLDLDCDGRWRREQA 180
 QY 181 ALDLYPDAGTDSGTFSSPNFATIPQDVTTEITSSPSHPANFSFYPRLKALPPIARVT 240
 Db 181 ALDLYPDAGTDSGTFSSPNFATIPQDVTTEITSSPSHPANFSFYPRLKALPPIARVT 240
 QY 241 LVRLRQSPRAFPAPVLPVPSRDNEIVDSASVPETPLDCEVSLWSWGLCGH 292
 Db 236 ----RGDTGAATAEPQGLHPSRPSQAQDNALVDSASVPETPLDCEVSLWSWGLCGH 291
 QY 293 CGRLGTSKTRVYRVOPANNGPCPELEEEACVDPNCV 331
 Db 292 CGRLGTSKTRVYRVOPANNGPCPELEEEACVDPNCV 330

RESULT 3

US-07-862-021B-12
 Sequence 12, Application US/07862021B
 Patent No. 5279966
 GENERAL INFORMATION:
 APPLICANT: Jessell, Thomas M
 APPLICANT: Klar, Avihu
 TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
 TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
 NUMBER OF SEQUENCES: 20
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham
 STREET: 30 Rockefeller Plaza
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10112
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/862,021B
 FILING DATE: 19920405
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 40028
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 977-9550
 TELEFAX: (212) 664-0525
 TELEX: 422523 COOP UI
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 802 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-862-021B-12

Query Match 26.5%; Score 464.5; DB 1; Length 802;
 Best Local Similarity 33.2%; Pred. No. 1.6e-37;
 Matches 111; Conservative 52; Mismatches 144; Indels 27; Gaps 10;
 9 ALGKALCALLLATLGAAGQPLGGESICSRALAKYSITFTGKWSQTAFFPKQYPLFRPPAQ 68
 Db 171 SLTKRICEQDSASEGVTDKP---TLDCACGCTAKYRLTFYGNWSEKTHPRDFF--RRNH 225
 QY 69 WSSLLGAHSDYSWMRKNOYVSNGLRDFAEERGEAWALMKET--EAAAGEALQSVHAFSAP 127
 Db 226 WSAIGSSHSKNYLWEYGGYASEGVKQVAELGSPVKMEEIQQQSDVLTIVIKAKAQPW 285
 QY 128 AVP--STGTSQSAELEVORRHSLVSFVVRIVPSDFVGVDSLDLDCDGR--RWREQAALDL 184
 Db 286 AWQPLNVRAPSAEFSVDRHRHLMSELTMLGSPDMNVNGLSAEDLCTKDCGWVQKVVDL 345
 QY 185 YPDAGTDSGTFSSPNFATIPQDVTTEITSSPSHPANFSFYPRLKALPPIARVTLVRL 244
 Db 346 IPMDAGTDSGVTVESPNKPTVQEKIRPLUSL--DHQSPFYDPGSGIKLVARVVLRI 403
 QY 245 RQSPRA--FIPP-----APVLPSPRNEIVDSASVPETPLDCEVSLWSWGLCGHCGRL 296
 Db 404 ARKEQCNFVDPNDIDIVADLAPEEKEE-----DDTPTCIYSNWSWPSACSSSTCEK 456
 QY 297 GTSKTRVYRVOPANNGPCPELEEEACVDPNC 330
 Db 457 GKRMRORMLKAO--LDLSVPCPDYQDFQPCMGPGC 489

Sequence 12, Application PC/TUS9303164
GENERAL INFORMATION:
APPLICANT: Jessell, Thomas M
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
NOVEL SECRETED PROTEIN, F-SPONDIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03164
FILING DATE: 19930402
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-03164-12

Query Match 26.5%; Score 464.5; DB 5; Length 802;
Best Local Similarity 33.2%; Pred. No. 1.6e-37;
Matches 111; Conservative 52; Mismatches 144; Indels 27; Gaps 10;
QY 9 ALGKALCALLATLGAAGQPLGGESICARALAKYSITFTGKWSQTAFFPKQYPLFRPPAQ 68
DB 171 SLTKRICEQDSASEGVTDKP---TLDCACGTAKYRLTFYGNWSEKTHPKDFP--RRTNH 225
QY 69 WSSLGAAHSDYSMMWRKNQYVNSGLRDFAEERGEAWALMKEI--EAGEALQSVHAFSAP 127
DB 226 WSAIIGSSHSHKNTLWEYGGYASEGVKQVAELGSPVKMEEEIRQSDDEVLTIVIKAKAOMP 285
QY 128 AVP--SGTGQTSAELEVORRHSLVSFVVRVPSDPWFVGVDSLDLDCGD--RWREQAALDL 184
DB 286 AQPLNVRAAPSAEFSVDRHRHLSFLTMGLSPDWNVGLSAEDLCTKDCGWQKVQVDL 345
QY 185 YPYDAGTDSGFTSSPNFATIPQDVTETITSSPSHPANSFYPRKALPIIARVTLVRL 244
DB 346 IPWDAGTDSGVTVESPNKPTVPQEKIRPLTSL--DHPQSPFYDPGSGIKLVARVVLRI 403
QY 245 ROSPRA--FIPP-----APVLPDRDNEIVDSASVPETPLDCEVSLWSGICGGHCGRL 296
DB 404 ARKGECNCFVDPNIDIVADLAPEEKEE-----DDTPETCIYSNWSFWSACSSTCEK 456
QY 297 GTKSRTRYVRVQPNANGSPCPPELEEEAECPDNC 330
DB 457 GKRMRMLKAO--LDLSVPCPDQDFQPCMGPGC 489

RESULT 5
PCT-US93-03164-12

US-08-313-288B-12
Sequence 12, Application US/08313288B
Patent No. 5750502
GENERAL INFORMATION:
APPLICANT: Jessell, Thomas M. and AviHu Klar
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
NOVEL SECRETED PROTEIN, F-SPONDIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,288B
FILING DATE: January 5, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0526
TELEX:
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-313-288B-12

Query Match 26.5%; Score 464.5; DB 1; Length 802;
Best Local Similarity 33.2%; Pred. No. 1.6e-37;
Matches 111; Conservative 52; Mismatches 144; Indels 27; Gaps 10;
QY 9 ALGKALCALLATLGAAGQPLGGESICARALAKYSITFTGKWSQTAFFPKQYPLFRPPAQ 68
DB 171 SLTKRICEQDSASEGVTDKP---TLDCACGTAKYRLTFYGNWSEKTHPKDFP--RRTNH 225
QY 69 WSSLGAAHSDYSMMWRKNQYVNSGLRDFAEERGEAWALMKEI--EAGEALQSVHAFSAP 127
DB 226 WSAIIGSSHSHKNTLWEYGGYASEGVKQVAELGSPVKMEEEIRQSDDEVLTIVIKAKAOMP 285
QY 128 AVP--SGTGQTSAELEVORRHSLVSFVVRVPSDPWFVGVDSLDLDCGD--RWREQAALDL 184
DB 286 AQPLNVRAAPSAEFSVDRHRHLSFLTMGLSPDWNVGLSAEDLCTKDCGWQKVQVDL 345
QY 185 YPYDAGTDSGFTSSPNFATIPQDVTETITSSPSHPANSFYPRKALPIIARVTLVRL 244
DB 346 IPWDAGTDSGVTVESPNKPTVPQEKIRPLTSL--DHPQSPFYDPGSGIKLVARVVLRI 403
QY 245 ROSPRA--FIPP-----APVLPDRDNEIVDSASVPETPLDCEVSLWSGICGGHCGRL 296
DB 404 ARKGECNCFVDPNIDIVADLAPEEKEE-----DDTPETCIYSNWSFWSACSSTCEK 456
QY 297 GTKSRTRYVRVQPNANGSPCPPELEEEAECPDNC 330
DB 457 GKRMRMLKAO--LDLSVPCPDQDFQPCMGPGC 489

RESULT 5
PCT-US93-03164-12

US-08-799-173A-7
; Sequence 7, Application US/08799173A
; Patent No. 5871969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, GREGG,
; APPLICANT: PATRICK J. DILLON
; TITLE OF INVENTION: HUMAN NEURONAL ATTACHMENT FACTOR-1
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08799,173A
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDERS A.
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-799-173A-7

Query Match 26.3%; Score 462.5; DB 2; Length 392;
Best Local Similarity 34.2%; Pred. No. 9e-38;
Matches 116; Conservative 49; Mismatches 137; Indels 37; Gaps 12;
QY 9 ALGKALCALLATLGAAGQPLGGESICARALAKYSITFTGKWSOTAFPKQYPLFRPPAQ 68
Db 26 SLTKLCEQDPTLDGVTDRPI---LDCCACGTAKYRTFYGNWSEKTHPKDYP--RRANH 80
69 WSSLGAAHSSDYSMWRKNQYVNSGLRDFAEERGEAWALMKEI-EAAGEALQSVHAFSAP 127
Db 81 WSAIIGGSHSKNYLVMEYGGYASGVKQVAELGSPVKMBEEIRQOSDEVLTIVIKAKAOWP 140
QY 128 -----AVPSGTGQTSAELEVQRHSLVSFVVRIVPSPDWFVGVDSLDLDCGD-RWRE 178
Db 141 SWQPVNVRAAP-----SAEFSVDRTRHLSFTLTMGSPDNNVGLSADLCTKEGQWQ 194
QY 179 QAALDLYPDAGTDSGFTFSSPNFATIPQDVTVEITSSPSHPANSFYPRKALPPIAR 238
Db 195 KVVQDLIPWDAGTDSGVTVESPNKPTIPQEKIRPLTSL--DHQPSFYDPGSGITQVAR 252
QY 239 VTLVRL-RQSPRAFIAPPVLPSPDRNETVDSASVPE-----TPLDCEVLSWSSWGLCGG 291
Db 253 VVIERIARKGECNIVPDNV---DDIVADLA--PEEKDEDDTPTETCIYSNWSPWACS 306
QY 292 HCGRLGTKSTRYRVVQPNNGSPCPPELEEEAECPVDC 330
Db 307 STCEKGRMRQRLKQAQ-LDLSVPCPDQDFQPCMGPGC 344

RESULT 7
US-07-862-021B-10
; Sequence 10, Application US/07862021B

Patent No. 5279966
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07862,021B
; FILING DATE: 19920405
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-021B-10

Query Match 26.3%; Score 462.5; DB 1; Length 807;
Best Local Similarity 34.2%; Pred. No. 2.6e-37;
Matches 116; Conservative 49; Mismatches 137; Indels 37; Gaps 12;
QY 9 ALGKALCALLATLGAAGQPLGGESICARALAKYSITFTGKWSOTAFPKQYPLFRPPAQ 68
Db 176 SLTKLCEQDPTLDGVTDRPI---LDCCACGTAKYRTFYGNWSEKTHPKDYP--RRANH 230
QY 69 WSSLGAAHSSDYSMWRKNQYVNSGLRDFAEERGEAWALMKEI-EAAGEALQSVHAFSAP 127
Db 231 WSAIIGGSHSKNYLVMEYGGYASGVKQVAELGSPVKMBEEIRQOSDEVLTIVIKAKAOWP 290
QY 128 -----AVPSGTGQTSAELEVQRHSLVSFVVRIVPSPDWFVGVDSLDLDCGD-RWRE 178
Db 291 SWQPVNVRAAP-----SAEFSVDRTRHLSFTLTMGSPDNNVGLSADLCTKEGQWQ 344
QY 179 QAALDLYPDAGTDSGFTFSSPNFATIPQDVTVEITSSPSHPANSFYPRKALPPIAR 238
Db 345 KVVQDLIPWDAGTDSGVTVESPNKPTIPQEKIRPLTSL--DHQPSFYDPGSGITQVAR 402
QY 239 VTLVRL-RQSPRAFIAPPVLPSPDRNETVDSASVPE-----TPLDCEVLSWSSWGLCGG 291
Db 403 VVIERIARKGECNIVPDNV---DDIVADLA--PEEKDEDDTPTETCIYSNWSPWACS 456
QY 292 HCGRLGTKSTRYRVVQPNNGSPCPPELEEEAECPVDC 330
Db 457 STCEKGRMRQRLKQAQ-LDLSVPCPDQDFQPCMGPGC 494

RESULT 8
US-08-313-288B-10
; Sequence 10, Application US/08313288B
; Patent No. 5750502

; GENERAL INFORMATION:
 ; APPLICANT: Jessell, Thomas M. and AviHu Klar
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
 ; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/313,288B
 ; FILING DATE: January 5, 1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0526
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; -LENGTH: 807 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-313-288B-10

Query Match 26.3%; Score 462.5; DB 1; Length 807;
 Best Local Similarity 34.2%; Pred. No. 2.6e-37;
 Matches 116; Conservative 49; Mismatches 137; Indels 37; Gaps 12;

QY 9 ALGKALCALLATLGAAGPLGSGESICARALAKYSITFTGKWSOTAFPKQYPLFRPPAQ 68
 DB 176 SLTKKLEQDPTLDGVTDRI---LDCCACGTAKYRLTFYGNWSEKTHPKDYP--RRANH 230
 QY 69 WSSLGAHSSDYSMRKNQVYNSGLRDFAEERGEAWALMKEI-EAAGEALQSVHAFSAP 127
 DB 231 WSAIIGGSHSKNYVLWEYGGYASEGVKQVAELGSPVKMEEIEIRQSDDEVLTIVIKAKAQP 290
 QY 128 -----AVPSGTGQTSAELEVQRHSLVSFVVRIVPSDPWFVGVDSLDLDCGD-RWRE 178
 DB 291 SWQPVNVRAP-----SAEFSVDRTRHLSFLTMGSPDNVNVGLSAEDLCTKECGWVQ 344
 QY 179 QAAALDLYPDAGTDSGFTSSPNFATIPQDTVTETITSSSPHPANSFYPRKALPPTAR 238
 DB 345 KVVQDLIPWDAGTDSGVTVESPNKPTIPQEKIRPLTSL--DHQSPFYDPEGGSITQVAR 402
 QY 239 VTLVRL-RQSPRAPIPPAPVLPSPRNEIVDSASVPE-----TPLDCEVLSWSSWGLCGG 291
 DB 403 VVIERIARKGECNIVPDNV-----DDIVADLA--PEEKDEDDTPTETCIYSNWSWPSACSS 456
 QY 292 HCGRLGTSRTRYVRVQPVANNNGSPCELEEEAEACVDPNC 330
 DB 457 STCEKGRMRQRMKLAQ-LDLSVPCPDPTQDFQPCMGPGC 494

RESULT 9
 PCT-US93-03164-10
 ; Sequence 10, Application PC/TUS9303164
 ; GENERAL INFORMATION:
 ; APPLICANT: Jessell, Thomas M
 ; APPLICANT: Klar, AviHu

; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
 ; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham
 ; STREET: 30 Rockefeller Plaza
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10112
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/03164
 ; FILING DATE: 19930402
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 40028
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 977-9550
 ; TELEFAX: (212) 664-0525
 ; TELEX: 422523 COOP UI
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; -LENGTH: 807 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US93-03164-10

Query Match 26.3%; Score 462.5; DB 5; Length 807;
 Best Local Similarity 34.2%; Pred. No. 2.6e-37;
 Matches 116; Conservative 49; Mismatches 137; Indels 37; Gaps 12;

QY 9 ALGKALCALLATLGAAGPLGSGESICARALAKYSITFTGKWSOTAFPKQYPLFRPPAQ 68
 DB 176 SLTKKLEQDPTLDGVTDRI---LDCCACGTAKYRLTFYGNWSEKTHPKDYP--RRANH 230
 QY 69 WSSLGAHSSDYSMRKNQVYNSGLRDFAEERGEAWALMKEI-EAAGEALQSVHAFSAP 127
 DB 231 WSAIIGGSHSKNYVLWEYGGYASEGVKQVAELGSPVKMEEIEIRQSDDEVLTIVIKAKAQP 290
 QY 128 -----AVPSGTGQTSAELEVQRHSLVSFVVRIVPSDPWFVGVDSLDLDCGD-RWRE 178
 DB 291 SWQPVNVRAP-----SAEFSVDRTRHLSFLTMGSPDNVNVGLSAEDLCTKECGWVQ 344
 QY 179 QAAALDLYPDAGTDSGFTSSPNFATIPQDTVTETITSSSPHPANSFYPRKALPPTAR 238
 DB 345 KVVQDLIPWDAGTDSGVTVESPNKPTIPQEKIRPLTSL--DHQSPFYDPEGGSITQVAR 402
 QY 239 VTLVRL-RQSPRAPIPPAPVLPSPRNEIVDSASVPE-----TPLDCEVLSWSSWGLCGG 291
 DB 403 VVIERIARKGECNIVPDNV-----DDIVADLA--PEEKDEDDTPTETCIYSNWSWPSACSS 456
 QY 292 HCGRLGTSRTRYVRVQPVANNNGSPCELEEEAEACVDPNC 330
 DB 457 STCEKGRMRQRMKLAQ-LDLSVPCPDPTQDFQPCMGPGC 494

RESULT 10
 US-09-03164-238-2
 ; Sequence 2, Application US/09022238
 ; Patent No. 617244
 ; GENERAL INFORMATION:

APPLICANT: Sytkowski, Arthur J. and Yang, Meiheng
TITLE OF INVENTION: A novel MPG-1 Gene that is differentially expressed in prosta
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/022,238
APPLICATION NUMBER: US/09/022,238
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/644,326
FILING DATE: 10-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Maravic-Magovcevic, Ivana
REGISTRATION NUMBER: P-43,338
REFERENCE/DOCKET NUMBER: NER-262CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-022-238-2

Query Match 23.9%; Score 420.5; DB 4; Length 132;
Best Local Similarity 65.4%; Pred. No. 2.8e-34;
Matches 83; Conservative 5; Mismatches 22; Indels 17; Gaps 2;
QY 200 PNFATIPQDVTETITSSPSHPANSFYYPRLKALPPIARVTLVRLRQ-----SPRAF 251
Db 15 POLRHHPADVTETITSSPSHPANSFYYPRLKACSH-----RQDGTGAAATEPQGL 65
QY 252 IPPAPVLSRONEIYDVSASVPETPLDCEVSLWSSWGLCGHCGRLGTSRTRYVRVOPAN 311
Db 66 HPSRSPQAQDNALVDSASVPETPLDCEVSLWSSWGLCGHCGRLGTSRTRYVRVOPAN 125
312 NGSPCPE 318
11111111
Db 126 NGSPCPE 132

RESULT 11
US-07-862-021B-14
Sequence 14, Application US/07862021B
Patent No. 5279566
GENERAL INFORMATION:
APPLICANT: Jessell, Thomas M
APPLICANT: Klar, Avihu
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,021B
FILING DATE: 19920405
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 568 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-862-021B-14

Query Match 18.8%; Score 331; DB 1; Length 568;
Best Local Similarity 33.1%; Pred. No. 1.7e-24;
Matches 87; Conservative 40; Mismatches 116; Indels 20; Gaps 9;
QY 78 SSDYSWRKNQYVNSGLRDFAEERGEAWLMKLEAAG-FALQSVHAFSAPVP--SGTG 134
Db 1 SGEYVLSMRQ-ASDGKQVAEIGSPVKMBEEIRQKGDVLTIVIKAKAQPAPQUNVRA 59
QY 135 QTSAELEVORRHSLVSFVVRIVPSPDVFGVDSLDLDCGD-RWREQAALDLYPDAGTDS 193
Db 60 APSAEFSVDRSRHLSFAMGSPDPNNVGLTSEDLTCTKCGVQKVQDLPWDAGTDS 119
QY 194 GTFTSSPNFATIPQDVTETITSSPSHPANSFYYPRLKALPPIARVTLVRLRQSPRAFI 252
Db 120 GVTVESPNKPTIPQDKIRPLTSLDHPQSPSMT----RGGPITPIARVVTIERARKGE--- 172
QY 253 PPAPVLSRONEIV-----DSASVPETPLDCEVSLWSSWGLCGHCGRLGTSRTRYVRV 307
Db 173 -QCNILIPDNVDDIADVLTVEEKDEDDTPTCTIYNSWPSWSSACSSATCDKGRMRORMLKA 231
QY 308 QPANNGSPCELEEEAEACVPDNC 330
Db 232 Q-LDLSVPCPDQTQDFECMGPGC 253

RESULT 12
PCT-US93-03164-14
Sequence 14, Application PC/TUS9303164
GENERAL INFORMATION:
APPLICANT: Jessell, Thomas M
APPLICANT: Klar, Avihu
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03164

FILING DATE: 19930402
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 568 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-93-03164-14

Query Match 18.8%; Score 331; DB 5; Length 568;
Best Local Similarity 33.1%; Pred. No. 1.7e-24;
Matches 87; Conservative 40; Mismatches 116; Indels 20; Gaps 9;

QY 78 SSDYSWRKNOYVSNGLRDFAEGRGEAWALMKEIEAAG-EALQSVHAFVSAPAVP--SGTG 134
DB 1 SCEYVLWNRQ-ASGVKQVAGLSFVWMBEIRQKGDVLTIVIKAKAQNFAWQPLNVRA 59
QY 135 QTSAEVORRHSLVSFVVRVPDVFVGVDSLDLDCGD-RWREQAALDLYPDAGTDS 193
DB 60 APSAEFSVDRSHLSFLAMGSPDPMVGLTSEDLCCKECGWQKVQVQDLPWDAGTDS 119
QY 194 GTFFSPNFATIPQDTVTEITS-SSPSHPANSFYPRKLALPPIARVTLVRLRQSPRAFI 252
DB 120 GVTYESPNKPTIPQDKIRPLTSLDHPQSPMT----RGPIIPIARVVIERIARKE--- 172
QY 253 PPAPVLPNRDNIV----DSASVPTPLDCEVLSWSSWGLCGHCGRLGTSKSTRYVRV 307
DB 173 -QCNIPONVDVLDVLTVEEKDEDDTETCIYSNWSWPSACSSATCDKGRMRQMLKA 231
QY 308 QPANNGSPCELEEEAECPVDC 330
DB 232 Q-LDLSVCPDQDFPCMGPGC 253

RESULT 13
US-8-799-173A-18
Sequence 18, Application US/08799173A
Patent No. 5871969
GENERAL INFORMATION:
APPLICANT: HASTINGS, GREGG,
APPLICANT: PATRICK J. DILLON
TITLE OF INVENTION: HUMAN NEURONAL ATTACHMENT FACTOR-1
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/799,173A
FILING DATE: 11-FEB-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:

NAME: BROOKES, ANDERS A.
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 53 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-799-173A-18

Query Match 17.6%; Score 309; DB 2; Length 53;
Best Local Similarity 100.0%; Pred. No. 8.6e-24;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 CEVLSWSSWGLCGHCGRLGTSKSTRYVRVQPANNGSPCELEEEAECPVDC 330
DB 1 CEVLSWSSWGLCGHCGRLGTSKSTRYVRVQPANNGSPCELEEEAECPVDC 53

RESULT 14
US-09-022-238-3
Sequence 3, Application US/09022238
Patent No. 6177244
GENERAL INFORMATION:
APPLICANT: SYKOWSKI, Arthur J. and Yang, Meiheng
TITLE OF INVENTION: A novel NPG-1 Gene that is differentially expressed in pros
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,238
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/644,326
FILING DATE: 10-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Maravic-Magovcevic, Ivana
REGISTRATION NUMBER: P-43,338
REFERENCE/DOCKET NUMBER: NER-262CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-09-022-238-3

Query Match 10.8%; Score 189; DB 4; Length 37;
Best Local Similarity 91.4%; Pred. No. 4.1e-12;
Matches 32; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 261 RNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGR 295
:||:|||||
Db 3 QDNALVDSASVPETPLDCEVSLWSSWGLCGGHCGR 37

RESULT 15

US-09-371-696-3
; Sequence 3, Application US/09371696
; Patent No. 6287777
; GENERAL INFORMATION:
; APPLICANT: Sytkowski, Arthur J.
; APPLICANT: Yang, Meiheng
; TITLE OF INVENTION: NOVEL NPC-1 GENE THAT IS DIFFERENTIALLY EXPRESSED IN PROSTATE
; TITLE OF INVENTION: TUMORS
; FILE REFERENCE: 01948/053002
; CURRENT APPLICATION NUMBER: US/09/371,696
; CURRENT FILING DATE: 1999-08-10
; EARLIER APPLICATION NUMBER: US 09/022,238
; EARLIER FILING DATE: 1998-02-11
; EARLIER APPLICATION NUMBER: US 08/644,326
; EARLIER FILING DATE: 1996-05-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-371-696-3

Query Match 10.3%; Score 181; DB 4; Length 37;
Best Local Similarity 88.6%; Pred. No. 2.6e-11;
Matches 31; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 261 RNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGR 295
:||:|||||
Db 3 QDNALVDSASVPETPLDCEVSLWSSWGLCGGHCGR 37

Search completed: December 10, 2001, 20:00:45
Job time: 2879 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2001, 13:38:51 ; Search time 75.12 Seconds
(without alignments)
326.388 Million cell updates/sec

Title: US-09-170-042A-2

Perfect score: 1756

Sequence: 1 MNPSPAAALGKALCALLIA.....NGSPCELEEEAFVPCVDPNCV 331

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1756	100.0	331	20	AAW92460 Human NAF-1 protei
2	1748	99.5	331	18	AAW23663 Human neuronal att
3	1747	99.5	331	20	AAW83328 Human mindin polyp
4	1746	99.4	331	22	AAW38872 Human polypeptide
5	1746	99.4	444	22	AAW40658 Human polypeptide
6	1744	99.3	331	21	AAV79561 Cancer specific pr
7	1742	99.2	331	22	AAW82472 Human extracellular
8	1741	99.1	331	19	AAW70589 Adhesion-modulatin
9	1741	99.1	331	20	AAW41721 Human PRO866 prote
10	1741	99.1	331	21	AAB44277 Human PRO866 (UNO4
11	1741	99.1	331	21	AAB33465 Human PRO866 prote

12	1741	99.1	331	21	AAV95349 Human PRO866 antit
13	1740	99.1	385	22	AAG75576 Human colon cancer
14	1329	75.7	249	21	AAB34753 Human secreted pro
15	1101.5	62.7	299	21	AAV73490 Human mindin-relat
16	583	33.2	290	20	AAW83329 F-spondin of chick
17	464.5	26.3	392	20	AAW92461 Rat F-spondin prot
18	462.5	26.3	807	14	AAW44241 F-spondin (FP5-9).
19	462.5	26.3	807	14	AAW44241 Rat F-spondin prot
20	462.5	26.3	807	20	AAV04264 Rat vascular smoot
21	462.5	26.3	807	21	AAB13074 Vascular endotheli
22	460.5	26.2	807	20	AAV04262 Human vascular smo
23	460.5	26.2	807	20	AAV04262 Bovine vascular sm
24	460.5	26.2	807	21	AAB13072 Vascular endotheli
25	460.5	26.2	807	21	AAB13073 Vascular endotheli
26	460.5	26.2	807	22	AAB74401 Human ovarian canc
27	420.5	23.9	132	20	AAV34113 Human NPG-1 protei
28	351	20.0	63	21	AAB27878 Protein fragment e
29	331	18.8	568	14	AAW44243 Xenopus F-spondin.
30	309	17.6	53	20	AAW92469 Human NAF-1 protei
31	277	15.8	105	21	AAB34693 Human secreted pro
32	250	14.2	63	21	AAB27877 Sequence homologou
33	155.5	8.9	182	21	AAB58354 Lung cancer associ
34	155.5	8.9	343	22	AAB93099 Human protein sequ
35	134	7.6	1551	22	AAM41081 Human polypeptide
36	134	7.6	1588	22	AAM39295 Human polypeptide
37	134	7.6	1588	22	AAW20155 Secreted protein S
38	128.5	7.3	50	20	AAW92468 Rat F-spondin Fsp-
39	126.5	7.2	264	22	AAB48109 Mouse TANGO 204 po
40	126.5	7.2	264	22	AAB48130 Mouse TANGO 204 va
41	124.5	7.1	264	22	AAB48131 Mouse TANGO 204 va
42	123.5	7.0	264	22	AAW92466 Rat F-spondin Fsp-
43	121.5	6.9	52	20	AAW92466 Human polypeptide
44	116.5	6.6	242	22	AAM42006 Human TANGO 204 po
45	116.5	6.6	264	22	AAB48105 Human TANGO 204 po

ALIGNMENTS

RESULT 1

AAW92460
ID AAW92460 standard; Protein; 331 AA.

XX AAW92460;

XX AC
XX DT 21-APR-1999 (first entry)

XX AC
XX DT Human NAF-1 protein.

XX AC
XX DT NAF-1; neuronal attachment factor-1; F-spondin analogue; treatment;
KW spinal cord injury; peripheral nerves damage; neural cell adhesion;
KW neurite extension; tumour cell metastasis; inhibitor; mobility; disease;
KW endothelial cell proliferation; tumour neovascularisation; haemostasis;
KW angiostatic agent; wound healing; diagnostic; neurotrophic; anticancer;
KW antimetastatic; anti-angiogenic; antimetastatic; human.

XX OS Homo sapiens.

Key	Location/Qualifiers
peptide	1..23
protein	/label= signal_peptide 24..331

XX US5871969-A.

XX 11-FEB-1999.

XX 12-FEB-1997; 97US-0799173.

XX 12-FEB-1996; 96US-0011519.

XX 12-FEB-1997; 97US-0799173.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Dillon PJ, Hastings G;
 XX WPI; 1999-166644/14.
 DR N-PSDB; AAX02019.
 XX Nucleic acid encoding human neuronal attachment factor-1 - used to
 PT e.g. treat spinal cord injuries, and inhibit tumor cell metastasis
 PT and neovascularization
 XX
 PS Claim 1a; Fig 1; 29pp; English.
 XX
 CC This sequence represents a novel human neuronal attachment factor-1,
 CC NAF-1. NAF-1 is an analogue of rat F-spondin and is used to treat spinal
 CC cord injuries and damage to peripheral nerves (by promoting neural cell
 CC adhesion and neurite extension), to inhibit tumor cell metastasis
 CC (particularly in small cell and breast carcinoma) and endothelial cell
 CC proliferation, adhesion and mobility, to reduce tumor neovascularisation,
 CC as angiostatic agents for tumor cells, to promote wound healing and to
 CC modulate hemostasis. NAF-1 may also be used to identify treatments and
 CC diagnoses for human disease. Fragments of the nucleic acid that do not
 CC encode NAF-1 peptides are useful as probes to isolate the NAF-1 gene, its
 CC allelic variants, full-length cDNA or related sequences, in chromosomal
 CC location by in situ hybridization or in Northern blotting, and as
 CC diagnostic probes or primers. The protein has neurotrophic, anticancer,
 CC antimetastatic, anti-angiogenic, antimetastatic and modulates adhesion,
 CC proliferation and mobility of cells.
 XX
 SQ Sequence 331 AA;

Query Match 100.0%; Score 1756; DB 20; Length 331;
 Best Local Similarity 100.0%; Pred. No. 7.1e-164;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MENPSAAALGKALCALLATLGAAGQPLGGESIC SARALAKYSTITFTGKWSQTAFPKQY 60
 Db 1 menpsaaalgakallatlgaagqplggesicsaralakyistftgkwsqtafpkqy 60
 QY 61 PLFRPPAQWSSLLGAHSSDYSMWRKNQVYNSGLRDFAEERGEAWALMKEIEAAGEALQSV 120
 Db 61 plfrppaqwssllgaahssdy smwrknqy vns glr dfaergeawalmkeieaagealqsv 120
 QY 121 HAVFSAVAPVSGTGTSAELEVRHSLVSVFVRIVPSDPVDFVSDLDLDCDGRWREQA 180
 Db 121 havfsavapvsgtgtsaelevrhslsvfvrivpsdpvdfvgsldldcdgrwreqa 180
 QY 181 ALDLYPYDAGTDSGFTFSSPNFATIPQDVTETITSSSPSHPANSFYPRKALPPIARVT 240
 Db 181 aldlypydagtdsgftfsspnfatipqdtvtetitssspsphpansfyprkalppiarvt 240
 QY 241 LVRLQSPRAFIPAPVLPSPRNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRIGTKS 300
 Db 241 lvrlqsprafipapvlp sprneivdsasvp etpldcevs lws swg lcgghcgrlg tks 300
 QY 301 RTRYVRVQPNNGSPCPPELEEAECVPDNCV 331
 Db 301 rtryvrqvpanngspcppeleeeaecvpdncv 331

RESULT 2
 AAW23663
 ID AAW23663 standard; Protein; 331 AA.
 XX
 AC AAW23663;
 XX
 DT 08-MAR-1998 (first entry)
 XX Human neuronal attachment factor-1.
 DE
 XX Neuronal attachment factor-1; NAF-1; human; cell-cell interaction;
 KW cell adhesion; spinal cord injury; peripheral nerve damage; tumour;
 KW metastasis; therapy; malaria; wound healing.

XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..23
 FT /label= sig_peptide
 FT 24..331
 FT /label= Mat_protein
 XX
 PN W09729189-A1.
 XX
 XX 14-AUG-1997.
 XX
 XX 12-FEB-1996; 96WO-US01857.
 XX
 XX 12-FEB-1996; 96WO-US01857.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Hastings GA;
 XX
 XX WPI; 1997-415344/38.
 DR N-PSDB; AAW78360.
 XX
 PT New isolated human neuronal attachment factor-1 - promotes cell-cell
 PT interaction and cell adhesion, used for treating e.g. spinal cord
 PT injury, peripheral nerve damage or tumours
 XX
 PS Claim 1; Page 42-43; 55pp; English.
 XX
 CC This protein comprises human neuronal attachment factor-1 (NAF-1),
 CC a novel protein that promotes cell-cell interaction and cell
 CC adhesion. Its amino acid sequence was deduced from a DNA molecule
 CC (see AAW78360) isolated from a cDNA library derived from human
 CC epithelial sarcoma. NAF-1 shows sequence homology to the rat
 CC F-spondin family. NAF-1 polypeptides, especially the mature
 CC protein, can be expressed in host cells. It can be used e.g. to
 CC treat spinal cord injuries or damage to peripheral nerves by
 CC promoting neural cell adhesion and neurite extension, to inhibit
 CC tumour cell metastases, inhibit endothelial cell proliferation,
 CC adhesion and motility, to decrease tumour neovascularisation, to
 CC be angiostatic for tumours, to promote wound healing, and to
 CC modulate haemostasis. It can also be used to screen for agonists
 CC and for antagonists useful e.g. in the prevention of malaria.
 XX
 SQ Sequence 331 AA;

Query Match 99.5%; Score 1748; DB 18; Length 331;
 Best Local Similarity 99.7%; Pred. No. 4.3e-163;
 Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MENPSAAALGKALCALLATLGAAGQPLGGESIC SARALAKYSTITFTGKWSQTAFPKQY 60
 Db 1 menpsaaalgakallatlgaagqplggesicsaralakyistftgkwsqtafpkqy 60
 QY 61 PLFRPPAQWSSLLGAHSSDYSMWRKNQVYNSGLRDFAEERGEAWALMKEIEAAGEALQSV 120
 Db 61 plfrppaqwssllgaahssdy smwrknqy vns glr dfaergeawalmkeieaagealqsv 120
 QY 121 HAVFSAVAPVSGTGTSAELEVRHSLVSVFVRIVPSDPVDFVSDLDLDCDGRWREQA 180
 Db 121 havfsavapvsgtgtsaelevrhslsvfvrivpsdpvdfvgsldldcdgrwreqa 180
 QY 181 ALDLYPYDAGTDSGFTFSSPNFATIPQDVTETITSSSPSHPANSFYPRKALPPIARVT 240
 Db 181 aldlypydagtdsgftfsspnfatipqdtvtetitssspsphpansfyprkalppiarvt 240
 QY 241 LVRLQSPRAFIPAPVLPSPRNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRIGTKS 300
 Db 241 lvrlqsprafipapvlp sprneivdsasvp etpldcevs lws swg lcgghcgrlg tks 300
 QY 301 RTRYVRVQPNNGSPCPPELEEAECVPDNCV 331

Db 301 rtryrvqpnnngspcpeleeeaeacvpdncv 331
|||||

RESULT 3

AAW83328
ID AAW83328 standard; Protein; 331 AA.

XX AAW83328;

DT 01-MAR-1999 (first entry)

DE Human mindin polypeptide.

XX Mindin; human; integrin ligand; angiogenic disease; cancer;
KW metastasis; chronic inflammatory disorder; rheumatoid arthritis;
KW atherosclerosis; macular degeneration, diabetic retinopathy;
KW restenosis; Alzheimer's disease; neural disorder;
KW tissue remodelling; therapy; diagnosis.

Homo sapiens.

PN WO9850073-A1.

PD 12-NOV-1998.

PF 07-MAY-1998; 98WO-US09476.

PR 09-MAY-1997; 97US-0046106.

PA (SMK) SMITHKLINE BEECHAM CORP.

PI Jonak ZL, Lane PA, Trulli SH, Tsui P;

XX WPI: 1999-034688/03.

DR N-PSDB; AAV72537.

PT A new polypeptide is the integrin ligand, human mindin - useful in
therapy and diagnostic assays for diseases such as those associated
with angiogenesis

PS Claim 1; Page 29; 39pp; English.

XX This is the amino acid sequence of human mindin, a novel integrin
ligand. It has homology to F-spondin, and may have a
multifunctional activity in normal and disease states. The
invention provides human mindin polypeptides and polynucleotides
(see also AAV72537-38) and methods for producing such polypeptides by
recombinant techniques. It also relates to methods for using such
polypeptides and polynucleotides in the treatment of angiogenic
diseases (cancer, cancer metastasis, chronic inflammatory disorders,
rheumatoid arthritis, atherosclerosis, macular degeneration,
diabetic retinopathy), restenosis, Alzheimer's disease, neural
disorders and tissue remodelling. The invention also relates to
methods for identifying agonists and antagonists/inhibitors and for
treating conditions associated with human mindin imbalance with the
identified compounds. Diagnostic assays for detecting diseases
associated with inappropriate human mindin activity or levels are
also claimed.

XX Sequence 331 AA;

Query Match 99.5%; Score 1747; DB 20; Length 331;
Best Local Similarity 99.4%; Pred. No. 5, 4e-163;
Matches 329; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MENPSPAALGKALCALLATLGAGQPLGSGSICARALAKYSTFTGKWQTAFFPKQY 60

Db 1 menpsaaalgkallatlgagqplgsgsicaralakytsftgnwstqafpkqy 60

QY 61 PLFRPPAQWSSLGAAHSSDYSWRKNQYVNSGLRDFAEERGEAWALMKIEAAGALQSV 120

|||||

Db 61 plfrppaqwssllgaahssdyswmrknqyvsnglrdfaergeawalikeiaagealqsv 120
QY 121 HAVESAPAVPSGTQQTSAELEVRHSLVSVFVVRIVPSDFWFGVDSLDLDCGDRWREQA 180
|||||
Db 121 havfsapavpsgtqtsaelevqrhslvsvfvrvivpspdwfvgsldldcdgdrwreqa 180
QY 181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTETITSSSPSPANSFYPRLKALPPTARVT 240
|||||
Db 181 aldlypydagtdsgftfsspnfatipqdtvtetitssspshpansfyprikalppiarvt 240
QY 241 LVRLRQSPRAFIIPAPVLPSPRDNVIDSASVPETPLDCEVSLWSSWGLCGHCGRLGTKS 300
|||||
Db 241 lvrlrqsprafipapvlpvpsrdneivdsasvpetpldcevsllwsswglcghcgrlgtks 300
QY 301 RTRYRVQPANNNGSPCPELEEEAEACVPDNCV 331
|||||
Db 301 rtryrvqpnnngspcpeleeeaeacvpdncv 331

RESULT 4

AAW38872

ID AAW38872 standard; Protein; 331 AA.

XX AC AAW38872;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 2017.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

DR N-PSDB; AAI58028.

XX Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -

PS Example 3; SEQ ID NO 2017; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
the encoded polypeptides (AAI38642-AAI42213) with nootropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and

DT 15-AUG-2000 (first entry)
 XX Cancer specific protein Prol08 useful as prostate cancer marker.
 DE Cancer specific gene: CSG; prostate cancer; diagnosis; monitoring;
 XX staging; imaging; metastasis; therapy; human; Prol08.
 KW Homo sapiens.
 XX
 OS WO200023108-A1.
 PN 27-APR-2000.
 XX 18-OCT-1999; 99WO-US23764.
 XX 19-OCT-1998; 98US-0104741.
 PR (DIAD-) DIADEXUS LLC.
 XX
 PA Ali SM, Sun Y, Salceda S, Recipon H, Caferkey R;
 XX WPI: 2000-339528/29.
 DR N-PSDB; AA295034.
 XX
 XX Diagnosing, detecting, staging, monitoring, imaging and treating
 PT cancers, especially useful for detecting prostate cancer comprises
 PT measuring changes in levels of cancer specific genes in cells, tissues
 PT and body fluids.
 XX
 PS Claim 6; Page 29-30; 35pp; English.
 XX
 CC The present sequence is that of the protein encoded by human cancer
 CC specific gene (CSG) prol08 (see AA295034). The CSG was identified
 CC using a suppression subtractive hybridization method. It is a
 CC diagnostic marker for prostate cancer. In 11 different healthy
 CC tissues examined, prol08 mRNA expression was highest in the ovary
 CC and uterus. Expression in healthy prostate was low. However,
 CC overexpression of CSG prol08 was observed in 13 of 13 prostate
 CC cancer tissues examined. CSGs comprising prol08, the encoded
 CC polypeptide, or an antibody that specifically binds CSG, are used
 CC in claimed methods for the diagnosis, detection, staging,
 CC monitoring, imaging and treatment of prostate cancer. The new
 CC methods provide earlier diagnosis for the presence and metastasis
 CC of prostate cancer, and can be used to determine if a cancer has
 CC metastasized, or to monitor the progress or stage of the disease
 CC when it has not metastasized.
 XX
 XX Sequence 331 AA:
 Query Match 99.3%; Score 1744; DB 21; Length 331;
 Best Local Similarity 99.4%; Pred. No. 1,le-162;
 Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MENPSAALGKALCALLATLGAAGQPLGGESICARALAKYSITFTCKWSQTAPPKOY 60
 DB 1 menpsaalgkallatlgaaqplggescsarapakysitftgkwsqtapfkdy 60
 QY 61 PLRPPAOWSSLLGAHSDSYSMWRKNQYVNSCLRDFAEGRGAWALMKEIEAAGEALQSV 120
 DB 61 plrppaqwssllgaahssdysmwrknqyvnsclrdfaegrgeawalmlkeieagealqsv 120
 QY 121 HAVFSAVAPSGTGTSAELEVRHSLVSFVVRIVPSPDWFVGVDSLDLCGDRWREQA 180
 DB 121 hevfsavapsgtgttsaelevrhlslvsfvvrivpdpdvwfvgvdsldlcgdrwreqa 180
 QY 181 ALDLYPDAGTSGFTTSPNFATIPQDRTVETITSSPSHPANSFYPRKALPPIARTV 240
 DB 181 aldlypydagtsgfttspnfatipqdvteitsspspsfyprikalpplartv 240
 QY 241 LVRLRSPRAPIPPAPVPLSRDNEIVDSASVPTPLDCEVSLWSSWGLCCGCGRLGTSK 300
 DB 241 lvrlrdsprafippavplpsrdnelvdsasvptpldcevsllwsswglcgcgrlgtks 300

QY 301 RTRVRYQPNANGSPCPLEEEAEACVDPNCV 331
 DB 301 rtrvryqpangspcpelaeacvdpncv 331
 RESULT 7
 AAB82472
 ID AAB82472 standard; Protein; 331 AA.
 XX AAB82472;
 XX 22-AUG-2001 (first entry)
 XX Human extracellular matrix protein RGL.
 DE RGL; human; extracellular matrix protein; prostate cancer;
 KW metastasis; tumour; benign prostatic hyperplasia; therapy;
 KW diagnosis; antitumour.
 XX Homo sapiens.
 OS
 XX
 PH Key Location/Qualifiers
 FT Domain 31..103
 FT /label= F-spondin_domain
 FT Domain 138..221
 FT /label= F-spondin_domain
 FT Domain 278..300
 FT /label= Thrombospondin_domain
 FT Peptide 28..46
 FT /label= Immunogenic_peptide
 FT /note= "specifically claimed in claim 25"
 FT Peptide 77..91
 FT /label= Immunogenic_peptide
 FT /note= "specifically claimed in claim 26"
 FT Peptide 188..210
 FT /label= Immunogenic_peptide
 FT /note= "specifically claimed in claim 27"
 FT Peptide 263..274
 FT /label= Immunogenic_peptide
 FT /note= "specifically claimed in claim 28"
 XX WO200144291-A2.
 XX 21-JUN-2001.
 XX 15-DEC-2000; 2000WO-US33901.
 XX 16-DEC-1999; 99US-0172370.
 XX 07-DEC-2000; 2000US-9966561.
 XX (SCHD) SCHERING AG.
 XX Harkins R, Parkes D, Parry G, Schneider DW, Steinbrecher R;
 XX WPI: 2001-398128/42.
 XX N-PSDB; AAF90566, AAF90567.
 XX Novel human extracellular matrix polypeptide, RGL, useful in research,
 XX diagnosis and treatment of metastasis such as prostate cancer
 PS Claim 1; Fig 2; 69pp; English.
 XX The present sequence is that of human RGL, a new homologue of the
 XX extracellular matrix protein family, which is expressed in prostate
 XX tissue and may be over-expressed in prostate tumours. It shows
 XX homology to mindin and f-spondin proteins. This protein family is
 XX distinguished by 2 conserved spondin domains near the N-terminus
 XX and at least 1 thrombospondin type 1 repeat at the C-terminus. The
 XX rgl gene (see AAF90566) was isolated from a database screening. The
 XX invention provides human RGL polypeptides (especially corresponding
 XX to amino acids 1-331, 21-331, 27-331, 28-46 and 77-91 of RGL),
 XX polynucleotides encoding them, methods for producing the

CC polypeptides, expression vectors, genetically engineered host cells
 CC for expression of the polypeptides, and antibodies which
 CC specifically bind RGI or its immunologically active fragment
 CC (especially a polypeptide comprising amino acid 28-46, 77-91,
 CC 188-210, or 263-274 of RGI). It also provides methods for utilizing
 CC the polynucleotides, polypeptides and antibodies in research, can
 CC diagnosis and therapeutic applications. Thus, the antibodies can
 CC be used for diagnosis and/or detection of RGI expression, or
 CC can be conjugated to a therapeutic agent, such as a cytotoxic
 CC agent, for administration to patient for treatment of diseases
 CC characterised by RGI activity or expression, such as prostate
 CC cancer.

XX Sequence 331 AA;

Query Match 99.2%; Score 1742; DB 22; Length 331;
 Best Local Similarity 99.4%; Pred. No. 1.7e-162;
 Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 MENPSPAAALGKALCALLLATLGAAGPLGGESIC SARALAKYSITFTGKWSQTAFPPKQY 60
 |||||
 1 menpspaaalgalcalllatlgaagqplggescsagapakysitftgkwsqtatpkqy 60

QY 61 PLFRPPAQWSSLLGAHSSDYSMWRKNQYVSNGLRDFAEERGEAWALMKEIEAAGEALQSV 120
 |||||
 Db 61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

QY 121 HAVFSAAPVPSGTGOTSARELEVORRHSLVSFVVRIVPSPDFVGVDSLDLDCGDRWRREQA 180
 |||||
 Db 121 havfsapavpsgtgotsaelevorrrhslvsfvvrivpspdwfvgvdsldldcgdrwrreqa 180

QY 181 ALDLYPYDAGTDSGFTFSSPNFATIPQDVTVEITSSSPSHPANSEYYPRLKALPPIARVT 240
 |||||
 Db 181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprrlkalppiarvt 240

QY 241 LVRLQSPRAFIPAPVLPSPRNEIVDSASVPETPLDCEVLSWSSWGLCGHCGRLGTGS 300
 |||||
 Db 241 lvrlqsprafipapvlpvpsrdneivdsasvpetpldcevsllswswgllcgghcgrlgtks 300

QY 301 RTRYVRVQPNANGSPCPELEEEAECPDNCV 331
 |||||
 Db 301 rtryvrqvpanngspcpeleeeaeecvpdncv 331

RESULT 8
 AAW70589
 ID AAW70589 standard; Protein; 331 AA.

AAW70589;

21-JAN-1999 (first entry)

Adhesion-modulating protein zsig25.

zsig25; adhesion-modulating protein; prostate cell; prostatic carcinoma;
 B-cell cancer; infertility; Wolf-Hirschhorn syndrome;
 chromosome 4 (p16.3).

Homo sapiens.

WO9845442-A2.

15-OCT-1998.

10-APR-1998; 98WO-US07117.

11-JUN-1997; 97US-0049288.

10-APR-1997; 97US-0043421.

(ZYMO) ZYMOGENETICS INC.

Sheppard PO;

XX WPI; 1998-557522/47.
 DR N-PSDB; AAV63241.
 XX
 PT New zsig25 protein and related nucleic acid, fusion proteins,
 PT vectors, transformed cells - and antibodies, involved in modulation
 PT of adhesion, used for diagnosis and treatment of prostatic and
 PT B-cell tumours, stimulation of haematopoietic cells, treatment of
 PT immune deficiency etc.
 XX
 PS Claim 7; Page 111; 161pp; English.

XX The present sequence represents a protein designated zsig25. The zsig25
 CC protein is an adhesion-modulating protein expressed at very high level
 CC in prostate cells. The protein is useful as a diagnostic marker for
 CC prostatic carcinoma and B-cell cancers, possibly also for infertility,
 CC and as a reagent for separating cancerous and non-cancerous cells.
 CC The products may also be used to diagnose or treat Wolf-Hirschhorn
 CC syndrome, associated with a deletion in the region of chromosome 4
 CC (p16.3) where the zsig25 gene is located.

XX Sequence 331 AA;

Query Match 99.1%; Score 1741; DB 19; Length 331;
 Best Local Similarity 99.1%; Pred. No. 2.1e-162;
 Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MENPSPAAALGKALCALLLATLGAAGPLGGESIC SARALAKYSITFTGKWSQTAFPPKQY 60
 |||||
 Db 1 menpspaaalgalcalllatlgaagqplggescsagapakysitftgkwsqtatpkqy 60

QY 61 PLFRPPAQWSSLLGAHSSDYSMWRKNQYVSNGLRDFAEERGEAWALMKEIEAAGEALQSV 120
 |||||
 Db 61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

QY 121 HAVFSAAPVPSGTGOTSARELEVORRHSLVSFVVRIVPSPDFVGVDSLDLDCGDRWRREQA 180
 |||||
 Db 121 havfsapavpsgtgotsaelevorrrhslvsfvvrivpspdwfvgvdsldldcgdrwrreqa 180

QY 181 ALDLYPYDAGTDSGFTFSSPNFATIPQDVTVEITSSSPSHPANSEYYPRLKALPPIARVT 240
 |||||
 Db 181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprrlkalppiarvt 240

QY 241 LVRLQSPRAFIPAPVLPSPRNEIVDSASVPETPLDCEVLSWSSWGLCGHCGRLGTGS 300
 |||||
 Db 241 lvrlqsprafipapvlpvpsrdneivdsasvpetpldcevsllswswgllcgghcgrlgtks 300

QY 301 RTRYVRVQPNANGSPCPELEEEAECPDNCV 331
 |||||
 Db 301 rtryvrqvpanngspcpeleeeaeecvpdncv 331

RESULT 9

AAW41721
 ID AAW41721 standard; Protein; 331 AA.

XX
 AC AAW41721;

XX 07-DEC-1999 (first entry)

Human PRO866 protein sequence.

Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
 probe; blood coagulation disorder; cancer; cellular adhesion disorder;
 secreted protein; transmembrane protein.

OS Homo sapiens.

XX WO9946281-A2.

XX 16-SEP-1999.

XX

PR	15-MAY-1998;	98US-0085579.
PR	15-MAY-1998;	98US-0085580.
PR	15-MAY-1998;	98US-0085582.
PR	15-MAY-1998;	98US-0085589.
PR	15-MAY-1998;	98US-0085697.
PR	15-MAY-1998;	98US-0085700.
PR	15-MAY-1998;	98US-0085704.
PR	18-MAY-1998;	98US-0086023.
PR	22-MAY-1998;	98US-0086392.
PR	22-MAY-1998;	98US-0086414.
PR	22-MAY-1998;	98US-0086430.
PR	22-MAY-1998;	98US-0086486.
PR	28-MAY-1998;	98US-0087098.
PR	28-MAY-1998;	98US-0087106.
PR	28-MAY-1998;	98US-0087208.
PR	30-JUL-1998;	98US-0094651.
PR	11-SEP-1998;	98US-0100038.
XX		
PA	(GETH) GENENTECH INC.	
XX		
PI	Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;	
XX		
DR	WPI: 1999-551358/46.	
DR	N-P5DB; AAZ34089.	
XX		
PT	New secreted and transmembrane polypeptides and their polynucleotides,	
PT	useful for treating blood coagulation disorders, cancers and cellular	
PT	adhesion disorders -	
XX		
PS	Claim 12; Fig 87; 530pp; English.	
XX		
CC	The present invention describes secreted and transmembrane polypeptides	
CC	and their polynucleotides. The nucleotide sequences are useful as	
CC	sources of probes, primers, for chromosome mapping, and for generation	
CC	of antisense sequences. They can also be used to create transgenic	
CC	animals. The proteins can be used to treat a variety of diseases and	
CC	disorders, depending on their function. Diseases that may be treated	
CC	include blood coagulation disorders, cancers and cellular adhesion	
CC	disorders. They may also be used to raise antibodies. AAZ33891 to	
CC	AAZ34338, and AA41685 to AA41774 represent polynucleotide and	
CC	polypeptide sequence given in the exemplification of the present	
CC	invention.	
XX		
SQ	Sequence 331 AA;	
Query Match 99.1%; Score 1741; DB 20; Length 331;		
Best Local Similarity 99.1%; Pred. No. 2.1e-162;		
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;		
QY	1	MENPSPAALGKALCALLATLGAAGQPLGGESICARALAKYSITFTGKWSQTAFPKQY 60
Db	1	menpspaaligkallatlgaaagqplggesicarsapakysifftgkwsqatpkay 60
QY	61	PLFRPPAOWSSLLGAHSSDYSMWRKNOYVSNGLRDFAEERGEAWALMKEIEAAGEALQSV 120
Db	61	plfrppaqwsllgaahssdywmrknoyvsnglrdfaergeawalmkeleaagealqsv 120
QY	121	HAFSAPAPVSGTQTSAAELEVQRHSLVSFVVRVPSDPWFVGVDSLDLDCDGRREQA 180
Db	121	hevfsapavpsgtqtsaelevqrhslvsfvvrivpsdpwfvgsldldcdgdrveqa 180
QY	481	ALDLYPYDAGTDSGFTSSPNFATIPQDTYTEITSSSPSPANSEYPRILKALPPIARVT 240
Db	181	aldlypydagtdsgftsspnfatipqdtvtetsspspansfypyrilkalppiarvt 240
QY	241	LVRLRQSPRAFIPAPVLPSPRDNIEVDSASVPETPLDCEVSLWSSWGLCGHCGRLGTSK 300
Db	241	lvrlrqsporafipapvlpdrdnieidsasvpetpldcevsllwsswglcgghcgrlgtks 300
QY	301	RTRYVRVQPPANNNGSPCEPELEEEAECPDNCV 331
Db	301	rtryvrqvpanngspcpeleeeaeecvpdncv 331

RESULT 10
 AAB44277
 ID AAB44277 standard; Protein; 331 AA.
 XX
 XX AAB44277;
 DT 08-FEB-2001 (first entry)
 DE Human PRO866 (UNQ435) protein sequence SEQ ID NO:236.
 XX
 KW Human; secreted protein; transmembrane protein; PRO; EST; cytosstatic;
 KW expressed sequence tag; detection; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO2000053756-A2.
 XX
 XX 14-SEP-2000.
 XX
 XX 18-FEB-2000; 2000WO-US04341.
 XX
 PR 08-MAR-1999; 99WO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0145698.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 DR WPI; 2000-611443/58.
 XX
 DR N-PSDB; AAC78507.
 XX
 PT Novel PRO polypeptides and polynucleotides used in detection methods,
 PT to target bioactive molecules to specific cells, and to modulate
 PT cellular activities .
 XX
 PS Claim 12; Fig 87; 636pp; English.
 XX
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 CC sequence tag) sequences which encode secreted or transmembrane PRO
 CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
 CC activity. The polynucleotides and polypeptides can be used for detecting
 CC the presence of PRO polypeptides in samples, for linking bioactive
 CC molecules to cells and for modulating biological activities of cells,
 CC using the polypeptides for specific targeting. The polypeptide targeting
 CC can be used to kill the target cells, e.g. for the treatment of cancers.
 CC The polypeptide pairs provide specific targeting of bioactive molecules
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
 CC the isolation of the PRO polynucleotide sequences.
 XX
 SQ Sequence 331 AA;

Query Match 99.1%; Score 1741; DB 21; Length 331;
 Best Local Similarity 99.1%; Pred. No. 2.1e-162;
 Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MENPSPAALGKALCALLLTLGAAGOPLGGSICARALAKYSITFTGKWSQTAPPKQY 60
 DB 1 menpspaalglkallatlgaagpplggsicsarapakysitftgkwsqtapfkpy 60
 QY 61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQVSNGLRDFAEERGEAWALMKEIEAAGCALQSV 120
 DB 61 plfrppaqwssllgaahssdyismwrknqyngldfaergeawalmkeieaagcalqsv 120
 QY 121 HAVFSAPAVPSGTGTSAELEVRHSLYSFVVRIVPSPDFVGVDSLDLDCDGRWRQQA 180
 DB 121 hevfsapavpsgtgtstaelevrrhslsvfvrivpspdwfvgvdsldlcdgrwrreqa 180
 QY 181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTETITSSSSHPSHANSFYPRKALPPARVT 240
 DB 181 aldlypydagtdsgftfsspnfatipqdtvtetitsssshpsansfyprkallppiarvt 240
 QY 241 LVRLQSPRAFIPAPVLPSPDRNEIVDSASVPETPLDCEVSLWSSWGLCGCGRLGRTKS 300
 DB 241 llvrlqsprafippavlpdrneivdsasvpetpldcevsllwsswglcgcgrlgrlgtks 300
 QY 301 RTRYVRVQPNANGSPCELEEEAECPDNCV 331
 DB 301 rtryvrvqpnnngspcpeleeeaeecvpdncv 331
 RESULT 11
 AAB333465
 ID AAB33465 standard; Protein; 331 AA.
 XX
 AC AAB33465;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Human PRO866 protein UNQ435 SEQ ID NO:258.
 XX
 KW Human; immune related disease; diagnosis; antinflammatory; cardiant;
 KW dermatological; antiarthritic; antirheumatic; immunosuppressive;
 KW haemostatic; antithyroid; antidiabetic; nootropic; neuroprotective;
 KW antianaemic; hepatotropic; virucide; antipsoriatic; antiallergic;
 KW antiasthmatic; systemic lupus erythematosus; rheumatoid arthritis;
 KW osteoarthritis; spondyloarthropathy; systemic sclerosis; sarcoidosis;
 KW idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;
 KW systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;
 KW autoimmune thrombocytopaenia; immune-mediated renal disease;
 KW demyelinating disease; hepatobiliary disease; Whipple's disease;
 KW inflammatory bowel disease; gluten-sensitive enteropathy;
 KW autoimmune disease; immune-mediated skin disease; allergic disease;
 KW immunological disease; transplantation associated disease;
 KW graft rejection; graft-versus-host-disease.
 OS Homo sapiens.
 XX
 XX WO2000053758-A2.
 XX
 PD 14-SEP-2000.
 XX
 XX 02-MAR-2000; 2000WO-US05841.
 XX
 PR 08-MAR-1999; 99WO-US05028.
 PR 10-MAR-1999; 99US-0123618.
 PR 12-MAR-1999; 99US-0123957.
 PR 23-MAR-1999; 99US-0125775.
 PR 12-APR-1999; 99US-0128849.
 PR 20-APR-1999; 99WO-US08615.
 PR 28-APR-1999; 99US-0131445.
 PR 04-MAY-1999; 99US-0132371.
 PR 14-MAY-1999; 99US-0134287.
 PR 02-JUN-1999; 99WO-US12252.

QY 241 LVLRLQSPRAFPAPVLPSPRDNEDVDSASVPETPLDCEVSLWSWGLCGGHCGRGLGTSK 300
 DB 295 lxrlqspafipapvpiprsdrneivdsasvpetpldcevsllswglcgghcgrlgtks 354
 QY 301 RTRVYRVQPNANGSPCPPELEAECPDNCV 331
 DB 355 rtrvyrvqpnangspcppeleaeacvpdncv 385

RESULT 14
 AAB34753
 ID AAB34753 standard; Protein; 249 AA.
 AC AAB34753;
 XX
 DT 26-JAN-2001 (first entry)
 XX
 DE Human secreted protein fragment encoded by DNA clone vp22 1.

Secreted protein; human; autoimmune disorder; multiple sclerosis; ulcer;
 systemic lupus erythematosus; rheumatoid arthritis; anaemia; stroke;
 haematopoiesis regulation; tissue regrowth; wound healing; haemophilia;
 Alzheimer's disease; Parkinson's disease; Shy-drager syndrome; cancer;
 contraceptive; infection; growth inhibition; hyperproliferative disorder;
 psoriasis.
 OS Homo sapiens.
 XX
 PN WO200055375-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 17-MAR-2000; 2000WO-US07285.
 XX
 PR 17-MAR-1999; 99US-0124808.
 PR 17-MAR-1999; 99US-0124916.
 PR 17-AUG-1999; 99US-0149639.
 PR 01-OCT-1999; 99US-0157247.
 PR 29-NOV-1999; 99US-0167824.
 PR 15-FEB-2000; 2000US-0182711.
 XX
 PA (ALPH-) ALPHAGENE INC.
 XX
 PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
 XX
 DR WPI: 2000-638211/61.
 N-PSDB: AAC59794.

Novel proteins and polypeptides useful for the treatment of e.g
 multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 cancer, Alzheimer's disease, Parkinson's disease, stroke, anemia and
 ulcers

Disclosure: Page 477-478; 493pp; English.

This invention relates to 59 human secreted proteins and the nucleotide
 sequences encoding them. Sequences AAC59788-C59846 and AAB34687-B34745
 represent the proteins and their encoding nucleotide sequences, and
 sequences AAB34746-B34771 represent fragments of the proteins. Probes
 for the DNA sequences are represented by sequences AAC59847-C59956. The
 proteins exhibit neuroprotective, dermatological, immunosuppressive,
 antiinflammatory, antianaemic, nootropic, antiparkinsonian,
 cerebroprotective, haemostatic, vulnary, cytostatic, antipsoriatic,
 antibacterial, virucide, and fungicide activity. The proteins and
 nucleotide sequences are useful as nutritional sources or supplements
 and in research. The proteins are useful for treating immune deficiency
 and disorders, which may be genetic or resulting from infections,
 autoimmune disorders such as multiple sclerosis, systemic lupus
 erythematosus, rheumatoid arthritis, and for treating myeloid or lymphoid
 cell deficiencies such as anaemias by regulating haematopoiesis. The
 proteins are also useful in compositions for bone, cartilage, tendon,
 ligament and/or nerve tissue growth or regeneration, for wound healing,

CC tissue repair and replacement and in the treatment of wounds, incisions
 CC and ulcers. Other uses include in the treatment of central and
 CC peripheral nervous system and neuropathies such as Alzheimer's and
 CC Parkinson's diseases and Shy-Drager syndrome, and mechanical and
 CC traumatic disorders, such as spinal cord disorders, head trauma and
 CC stroke. The proteins may also be used as a contraceptive, and for
 CC treating coagulation disorders such as haemophilias. The protein and
 CC nucleotide sequences with cadherin activity are useful for treating
 CC cancer. Other uses for the protein include for inhibiting the growth,
 CC infection or function of, or killing, infectious agents such as bacteria,
 CC virus, fungi and other parasites, for effecting bodily characteristics
 CC such as height, weight, hair colour, effecting biorhythms or cardiac
 CC cycles or rhythms, effecting metabolism, catabolism, anabolism,
 CC processing, utilization, storage or elimination of dietary fat, lipid,
 CC protein, carbohydrate, vitamins, minerals, cofactors, effecting
 CC behavioural characteristics, providing analgesic effects and for treating
 CC hyperproliferative disorders such as psoriasis.
 XX
 SQ S-Sequence 249 AA;

Query Match 75.7%; Score 1329; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred. No. 3.8e-122;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 MWRKNQYVSNGLRDFAEERGEAWALMKIEAAGEALQSVHAFVSAPVPSGTGQTSAELEV 142
 DB 1 mwrknqyvsnglrdfaergeawalmkieaagealqsvhafsapvpsgtgtsaelev 60

QY 143 QRRHSLVSFVVRVPSDPWFVGVDSLDLCGDRWRQOALDLYPDAGTDSGTFSSPNF 202
 DB 61 qrrhslvsfvvrivpsdpwfvgvdsldlcdgdrwrqoalldlypydagtsgtfsspnf 120

QY 203 ATIPQDVTVTETSSPSHPANSEFYPRKALPPIARTVTLVRLRQSPRAFPAPVLPSPRD 262
 DB 121 atipqdvteitsspsphansfyprkappaiparvtrlvrlrqspafipapvlpssrd 180

QY 263 NEIVDSASVPETPLDCEVSLWSWGLCGGHCGRGLGTSKTRVYRVQPNANGSPCPPELEE 322
 DB 181 neivdsasvpetpldcevsllswswglcgghcgrlgtksrtrvyrvqpnangspcppelee 240

QY 323 AECVPDNCV 331
 DB 241 aecvpdncv 249

RESULT 15
 AAY73430
 ID AAY73490 standard; Protein; 299 AA.
 AC AAY73490;
 XX
 DT 29-FEB-2000 (first entry)
 XX
 DE Human secreted protein clone yk143_1 protein sequence SEQ ID NO:202.
 XX
 KW Human; secreted protein; immunostimulatory; haemostatic; cytokine;
 KW proliferative; differentiative; chemotactic; chemokinetic; vaccine;
 KW thrombolytic; antiinflammatory; cytostatic; immunosuppressive;
 KW gene therapy.
 XX
 OS Homo sapiens.
 PN WO9958642-A2.
 XX
 PD 18-NOV-1999.
 XX
 PF 14-MAY-1999; 99WO-US10843.
 PR 14-MAY-1998; 98US-0085472.
 PR 17-AUG-1998; 98US-0096824.
 PR 11-SEP-1998; 98US-0099843.
 PR 11-SEP-1998; 98US-0099950.

PR 15-SEP-1998; 98US-0100424.
PR 29-SEP-1998; 98US-0102329.
PR 09-OCT-1998; 98US-0103615.
PR 11-DEC-1998; 98US-0111799.
PR 14-DEC-1998; 98US-0112159.
PR 31-DEC-1998; 98US-0114415.
PR 10-FEB-1999; 98US-0248059.
PR 06-APR-1999; 99US-0287150.
PR 13-MAY-1999; 99US-0311021.
XX
PA (GENY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ;
XX
DR WPI: 2000-053095/04.
DR N-PSDB; AAZ52575.
XX
PT Novel polynucleotides and proteins having biological activities which
PT make them suitable for treating, preventing or ameliorating medical
PT conditions in humans or animals -
XX
Claim 211; Page 710-711; 730pp; English.
XX
CC The present invention describes human secreted proteins encoded by
CC polynucleotides obtained from adult testes, foetal brain, adult brain,
CC brain (foetal and adult), foetal kidney, adult spleen, and adult thymus
CC cDNA libraries. The polynucleotides and proteins are predicted to have
CC biological activities which would make them suitable for treating,
CC preventing or ameliorating medical conditions in humans and animals.
CC Suggested activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccines) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy. Therapeutic compositions
CC are also presently valuable for veterinary applications. AAZ52475 to
CC AAZ52581 encode human secreted proteins, and AAY73390 to AAY73500
CC represent human secreted proteins, given in the present invention.
XX
SQ Sequence 299 AA;

Query Match 62.7%; Score 1101.5; DB 21; Length 299;
Best Local Similarity 90.3%; Pred. No. 1e-99;
Matches 214; Conservative 5; Mismatches 15; Indels 3; Gaps 2;

83 MWRKNQVYVNGLRDFAERGEAWALMKETEAGEALQSVHAFVPSAPVPSGTGQTSAELEV 142
1 mwrknqvysnglrdfaergeawalmkeieagealqsvhvfapvpsgtgtsaelev 60

143 QRRHSLVSFVVRIVPSPDWFVGVDSLDLDCGDRWREQAALDLYPYDAGTDSGFTFSSPNF 202
qrrhslvsfvvrivpspdwfvgvnsldlcdgdwrreqaaldlypydagtdsggftfsspnf 120

61 qrrhslvsfvvrivpspdwfvgvnsldlcdgdwrreqaaldlypydagtdsggftfsspnf 120

203 ATIPQDVTETSSSPSPANSFYPRKALPPIARVTLVRLQSPRAFIPPAVPLESRD 262
atipqdvteitssspshpansfyprkallppiarvtrlvrlqsprafippavplpsrd 180

263 NEIVDSASVPETPLDCEVLSWSSGCGHGRGLGKTRFYRVQPA--NNGSPCP 317
neivdsasvpetpldcevlswswgicggplweardqeqds-lrpgpqdgtapap 236

181 neivdsasvpetpldcevlswswgicggplweardqeqds-lrpgpqdgtapap 236

Search completed: December 10, 2001, 19:46:03
Job time: 22032 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2001, 19:17:38 ; Search time 16.91 Seconds
(without alignments)
1491.057 Million cell updates

```

US-09-170-042A-2
1756
perfect score:
Sequence: 1 MENPSPAAALGKALCALLA.....NGSPCELEEEAECPDNCV 331

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

```

Searched: 219241 seqs, 76174552 residues

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	462.5	26.3	807	2	A38152
2	448.5	25.5	803	2	F-spondin - rat
3	406	23.1	805	2	F-spondin precursor
4	127.5	7.3	741	2	hypothetical prote
5	117	6.7	772	2	probable transcrip
6	108	6.2	123	2	transcription fact
7	108	6.2	742	2	TRAP-C2 protein -
8	107	6.1	808	2	transcription fact
9	104.5	6.0	440	2	A49672
10	102	5.9	1251	2	phospholipase D (E
11	104	5.8	483	1	hypothetical prote
12	100.5	5.7	810	2	latent transformin
13	100.5	5.7	2957	2	minor coat protein
14	99.5	5.7	424	2	hypothetical prote
15	93.5	5.7	590	2	hypothetical prote
16	99.5	5.7	724	2	complement compone
17	99.5	5.7	812	2	antigen Em100 - Ei
18	97.5	5.6	808	2	phospholipase D (E
19	97	5.5	534	2	phospholipase D (E
20	97	5.5	1360	2	hypothetical prote
21	97	5.5	1584	2	hypothetical prote
22	96.5	5.5	598	2	brain-specific ang
23	96.5	5.5	712	2	beta-galactosidase
24	96.5	5.5	812	2	immunodominant mic
25	96	5.5	591	1	probable phospholi
26	95	5.4	1070	2	complement c8 beta
27	94	5.4	809	2	cellulase (EC 3.2.
28	93	5.3	3027	2	phospholipase D (E
29	92.5	5.3	912	2	polyprotein - pars
30	92.5	5.3	45423	2	brevican precursor

ALIGNMENTS

RESULT 1

A38152

F-spondin - rat

C-Species: Rattus norvegicus (Norway rat)

C-Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 17-Nov-2000

C-Accession: A38152

R-Klar, A.; Baldassare, M.; Jessell, T.M.

Cell 59, 95-110, 1992

A-Title: F-spondin: a gene expressed at high levels in the floor plate encodes a secretory glycoprotein

A-Reference number: A38152; MUID:92208952.

A-Accession: A38152

A-Status: preliminary

A-Molecule type: mRNA

A-Residues: 1-807 <LA>

A-Cross-references: GB:M88469; NID:g204176; PID:AAA41174.1; PID:g204177

A-Experimental source: embryo floor plate

A-Note: sequence extracted from NCBI backbone (NCBIN:90877, NCBIP:90878)

C-Superfamily: F-spondin; thrombospondin type 1 repeat homology

F:441-495/Domain: thrombospondin type 1 repeat homology <THR2>

F:500-555/Domain: thrombospondin type 1 repeat homology <THR3>

F:557-611/Domain: thrombospondin type 1 repeat homology <THR4>

F:613-666/Domain: thrombospondin type 1 repeat homology <THR1>

F:667-721/Domain: thrombospondin type 1 repeat homology <THR5>

F:753-807/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match	26.38;	Score	462.5;	DB 2;	Length	807;
Best Local Similarity	34.28;	Prod. No.	2.9e-29;			
Matches 116;	Conservative	49;	Mismatches	137;	Indels	37;
					Gaps	12;

QY	3	ALGKALCALLATLGAAGQPLGGESICSRALAKYITTTGKWSOTAPFKQYPLFRPPAQ	68
DB	176	SLTKKLCQEDTLQGVTRPI---LDCACCTAKYRLTFYGNWSEKTHPKDP--RRANH	230
QY	69	WSSLGAAHSDYSNMRKNQYVSNGLRDFARGEAALMKEI-ERAGEALQSVHAFSP	127
DB	231	WSAIIGGSHSNYLMWEYGGYASGVQKVAELGSPVKMEETRQOSDEVLTVIKAKAQP	290
QY	28	-----AVPSGTCQSAELEVORRHLSYFVVRIVPSPDMFVGVDLSLDLDCGD-RWRE	178
DB	291	SNQPVNVRAP-----SAEFSVDTRHLSFLTMGMPGDMNVGLSAEDLCKEKGWQ	344
QY	179	QAALDLYPYDAGTSGDTFFSPNFATIPQDTVEITSSPSHPANSFYPRLKALPPIAR	238
DB	345	KVVQDLIPWDAGTDSGVTYESPKNPTIPQEKIRPLTSL--DHPQSPFYDPEGSSITQVAR	402
QY	239	VTLVRL-RQSPAFITPPAPVLPISRDNEIVDSASVPE-----TPLDCEVSLWSMGLCGG	291
DB	403	VVIERIARKGECQIVPNV----DDIVADLA--PEEKDEDTPTETCIYSNNPSWACS	456
QY	292	HCGRLGTCSRYYRVQVPANNGSPCELEEEACVDPNC	330

QY 230 LKALPPIARVTLVLRQSPRAFIP-----PAPVLPSRDNEIVDSASVPETP----- 275
DB 353 VESLPVASSSTLLPLVPSNSISLSTFGSTNLACGFFPFSQNLGTANDTSGPELDPGLGL 412
QY 276 -----LDCEVSLWSWGLCGHCGRLGTSKTRVYRVQPANNGSPCPPELEEE 322
DB 413 LDEAMLD-EISLMD-----LATEEGNPVQASQLEEE 443

RESULT 5
A55004
transcription factor TFC11 - human
N:Alternate names: LCR-F1 protein
C:Species: Homo sapiens (man)
C:Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 05-Nov-1999
C:Accession: A55004; S48097
R:Luna, L.; Johnsen, O.; Skartlien, A.H.; Pedetour, F.; Turc-Carel, C.; Prydz, H.; Kols
Genomics 22, 553-562, 1994
Title: Molecular cloning of a putative novel human bZIP transcription factor on chromo
some 17p11.2.
Reference number: A55004; MUID:95095252
Accession: A55004
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-772 <LUN>
A:Cross-references: GB:X77366; NID:g541677; PIDN:CAAS4555.1; PID:g541678
R:Caterina, J.J.; Donze, D.; Sun, C.W.; Ciavatta, D.J.; Townes, T.M.
Nucleic Acids Res. 22, 2383-2391, 1994
A:Title: Cloning and functional characterization of LCR-F1: a bZIP transcription factor
A:Reference number: S48097; MUID:94310069
Accession: S48097
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 326-772 <CAT>
C:Genetics:
A:Gene: GDB:TCF11
A:Cross-references: GDB:293921; OMIM:600115
A:Map position: 17q22-17q22
C:Superfamily: fos/jun DNA-binding domain homology
C:Keywords: DNA binding; leucine zipper; transcription factor
F:648-689/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 6.7%; Score 117; DB 2; Length 772;
Best Local Similarity 23.3%; Pred. No. 0.16;
Matches 78; Conservative 35; Mismatches 110; Indels 112; Gaps 17;
19 LATLGAAGOPLGGSICSRALAKYSITFTG-KWSQTAFPKQYPLFRPPAQWSSLLGAH 77
153 LEDLGAVAPPVSGD-----LTKEIDILDLWRQDID-----LGAGR 189
QY 78 S-SDYSWMRKQYVNSGLRDPFAER-----GE-AWALMKEIEAAGALQSVHAVFSAPVPS 131
DB 190 EVFDYSHRQKQDVEKELRDGGEQDTWAGEAEALARNLLVDGETGESFPA-----QVPS 244
QY 132 GTGQTSAAELVQRR-----HSLVSFVVRIVPSPD-----WFVGV 165
DB 245 GEDQTALSLEECURLLEATCPGAEAPADISSITEAVPSEPPALQNLLSPLLTGT 304
QY 166 DS-LDLCDDRWRE-QAALDLPYDAGTD-SGFTSSPNFATIPQDVTETITSSPSHPA 222
DB 305 ESPFDL--EQWQDLMSIMQAEVNTSASEILYSAP-----PGDPLSTNYSAPNTPI 357
QY 223 NS-----FYVPRKALPPIARVTLVLRQSPRA-----F 251
DB 358 QNWSLHQASLGCSCQDFLLFPVESLPVASSSTLLPLAPSNSISLSTFGSTNLTGLF 417
QY 252 IPP-----APVLSRDNEIVDSASVPETPL 276
DB 418 FPPQLNGTANDTAGPELDPGLGLDEAMLEISL 452

RESULT 6

S49108
TRAP-C2 protein - Cryptosporidium parvum (fragment)
C:Species: Cryptosporidium parvum
C:Date: 01-Feb-1995 #sequence_revision 12-May-1995 #text_change 09-Sep-1997
C:Accession: S49108
R:Spano, F.S.; Ranucci, L.R.; Catteruccia, F.C.; Saccheo, S.S.; Crisanti, A.C.
submitted to the EMBL Data Library, January 1994
A:Description: Thrombospondin related protein in Cryptosporidium.
A:Reference number: S49108
A:Accession: S49108
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-123 <SPA>
A:Cross-references: EMBL:X77586; NID:g509278; PID:g509279
C:Superfamily: thrombospondin type 1 repeat homology
F:6-61/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 6.2%; Score 108; DB 2; Length 123;
Best Local Similarity 40.4%; Pred. No. 0.089;
Matches 23; Conservative 6; Mismatches 26; Indels 2; Gaps 2;
QY 275 PLDCEVSLWSWGLCGHCGRLGTSKTRVYRVQPAN-NGSPCPPELEEEAECVPDNC 330
DB 5 PLSCVTSEWGNWSEKSLTCG-IGHQMRERSVIRKAPQDNLFQCPETRIQECIQDTC 60

RESULT 7
A49672
transcription factor Nr1f - human
N:Alternate names: Basic leucine zipper transcription factor; NF-E2-related factor 1
C:Species: Homo sapiens (man)
C:Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 16-Dec-1998
C:Accession: A49672
R:Chan, J.Y.; Han, X.L.; Kan, Y.W.
Proc. Natl. Acad. Sci. U.S.A. 90, 11371-11375, 1993
A:Title: Cloning of Nr1f, an NF-E2-related transcription factor, by genetic selection
A:Reference number: A49672; MUID:94068605
Accession: A49672
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-742 <CHA>
A:Experimental source: erythro leukemia cell line K562
A:Note: sequence extracted from NCBI backbone (NCBI:140521, NCBI:140522)
C:Superfamily: fos/jun DNA-binding domain homology
F:618-659/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 6.2%; Score 108; DB 2; Length 742;
Best Local Similarity 23.2%; Pred. No. 0.78;
Matches 73; Conservative 38; Mismatches 103; Indels 100; Gaps 16;
QY 19 LATLGAAGOPLGGSICSRALAKYSITFTG-KWSQTAFPKQYPLFRPPAQWSSLLGAH 77
DB 153 LEDLGAVAPPVSGD-----LTKEIDILDLWRQDID-----LGAGR 189
QY 78 S-SDYSWMRKQYVNSGLRDPFAER-----GE-AWALMKEIEAAGALQSVHAVFS 126
DB 190 EVFDYSHRQKQDVEKELRDGGEQDTWAGEAEALARNLLVDGETGESFPAQFADISSI 249
QY 127 -PAYPSGTGQTSAAELVQRRHSLVSFVVRIVPSPDWFVGVDLDCDGRWRE-QAALDL 184
DB 250 TEAVPSEPPALQ-----NNLSPLLTGTSP-----FDL--EQWQDLMSIM 293
QY 185 YPYDAGTD-SGFTSSPNFATIPQDVTETITSSPSHPANS-----FY 226
DB 294 QAMEVNTSASEILYSAP-----PGDPLSTNYSAPNTPIQNVSLHQASLGCSCQDFLLF 348
QY 227 YPRKALPPIARVTLVLRQSPRA-----FIPP-----APVLSRD 262
DB 349 SPEVESLPVASSSTLLPLAPSNSISLSTFGSTNLTGLTFPPQLNGTANDTAGPELDP 408
QY 263 NEIVDSASVPETPL 276

Db 409 GGLDEAMLDEISL 422

RESULT 8

T10171
 phospholipase D (EC 3.1.4.4) - castor bean
 N:Alternate names: choline phosphatase; lecithinase D; lipophospholipase II
 C:Species: Ricinus communis (castor bean)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 13-Aug-1999
 C:Accession: T10171; T10118; A54850
 R:Xu, L.; Zheng, L.; Coughlan, S.J.; Wang, X.
 Plant Mol. Biol. 32, 767-771, 1996
 A>Title: Structure and analysis of phospholipase D gene from Ricinus communis L.

A:Reference number: Z16972; MUID:97134969
A:Accession: T10171
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-808 <XUL>
A:Cross-references: EMBL:U72693; NID:g1698843; PTD:g1698844
Experimental source: cultivar Hale; tissue_type leaf
Submitted to the EMBL Data Library, July 1996

A:Reference number: Z16957
A:Accession: T10118
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-267, 'L', 269-808 <WAN>
A:Cross-references: EMBL:I33686; NID:g1438075
A:Experimental source: strain Hale; tissue_type endosperm
R:Wang, X.; Xu, L.; Zheng, L.
J. Biol. Chem. 269. 20312-20317. 1994

A:Title: Cloning and expression of phosphatidylcholine-hydrolyzing phospholipase
A:Reference number: A54850; MUID:94327597
A:Accession: A54850
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-101,'p','103-267','L','269-447','HLLLC','453-456','LRHGMSCLDPLAVELHLYSL'
A:Cross-references: GB:L33686
A:Note: the authors translated the codon AAG for residue 535 as Leu
C:Genetics

C:denatetics.
A:introns: 31/3: 665/1
C:keywords: phosphoric diester hydrolase

Query Match	6.1%	Score 107;	DB 2;	Length 808;
Best Local Similarity	24.6%	Pred. No. 1;		
Matches 59;	Conservative 24;	Mismatches 81;	Indels 76;	Gaps 14;

DB 303 VHCVL-CPRNPDDGSGFVODLIQISTMFTTHQIKVIVVDSAMPNGDSQRRRIYSFVG -GLD 359
 120 VHAVESAPAVPSGTCGTSAELEVQ---RHSLYSFVFWIRVSPD-----WFGVQSLD 169
 170 LCDGDRREQAALDLYPYDAGTDCGF-----TFSPNFATTPQDVTVEITSSPSH 220
 360 LCDG-----RYDSPHSLFRLDLSAHHDDFHQPNFAG-----ASIEKGGPRE 401
 221 PANSFYYPRLKALPPA-----RVTLVLRQSPRAFIPPAPVLPSRONE 264

402	PWHDI-HSRLEG--PIANDVLFNFQRWRKGGKDLLQLQLEVEDVIIPSPVMYPDDFE	458
265	I-----VDSASV---PEPLDCEVLSWSGLCGHCGRLGTSKTRIVR-VQPA	312
459	ANNVLFNSIDGGAAGFPETPEDAPEA-----GLVSGKDNIIDRSIQDAYTHAIRRAKN	513
RESULT	9	

T24232
hypothetical protein RL7.3 - Caenorhabditis elegans
;Species: Caenorhabditis elegans
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
;Accession: T24232
R:Barlow, K.

submitted to the EMBL Data Library, March 1997
 Accession number: Z10860

A: Accession number: Z13600
 A: Accession: T24232
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-440 <WIL>
 A: Cross-references: EMBL:Z92809; PIDN:CAB07269.1; G
 A: Experimental source: clone R17
 C: Genetics:
 A: Gene: CRSP:R17.3
 A: Map position: 3
 A: Introns: 36/3; 89/3; 179/2; 241/1; 306/1; 346/2
 C: Superfamily: Caenorhabditis elegans hypothetical

Query Match 6.0%; Score 104.5; DB 2; Length 440;
Best Local Similarity 25.0%; Pred. NO. 0.8;
Matches 44; Conservative 20; Mismatches 43; Indels 69; Gaps 11;

Qy 206 PQDVT-EITSSSPHANSFYPRKAL-----PPIARVTLVRLRQSPR----- 249

131 PRDVFVQATDITMDPA-----DRIKTI TECTYKQDQMDP----- 170

QY	250	-----AFIP----	PAPVL-	-----PSRNEIVDSASYPE-----	273
		:	:	: : :	

Db 180 YSLSKFMPLKVTSTPLLYEENRVQPANNLYLES-SISECYCDEHCVTGLGCCSDYTFV 238

QY 274 -TPLDCEVSLWSWGLC---GGHCGRGLGTKSKTRYYRVQPANNNGSPCPLEEEAEAC 325

Db 239 C P P R D C V L T D W D S W T Q C T A D N G T C G - I G T Q K R L R H V T Q H A E R G G A C E P L K E M R T C 293

RESULT 10
A57293
latent transforming growth factor beta-binding protein 3 precursor - mouse
N-Alternate names: mitosis-inhibitory protein

C;Accession: A57293; A60487
C;Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 11-Jan-2000
C;Species: Mus musculus (house mouse)
C;Synonyms: Mitosis inhibitory peptide

J. Biol. Chem. 270, 10147-10160, 1995

A;Accession: A57293
A;Status: preliminary
A;Molecule type: mRNA

A;Residues: 1-1251 <YIN>
A;Cross-references: GB:L40459
R;Reichert, K.L.; Paulsen, J.E.; Elgjo, K.
Virchows Arch. B Cell Pathol 59 137-142 1990

A;Title: Isolation of a growth and mitosis inhibitory peptide from mouse L-cells.
A;Reference number: A60487
A;Accession: A60487

A: Molecule type: protein
A: Residues: 65-69 <REI>
C: Comment: The molecular source of this pentapeptide has not been shown but
C: Genetics:

A: Gene: *Ltbp-3*
C: Superfamily: unassigned EGF-related proteins; EGF homology
C: Keywords: liver; pyroglutamic acid

F:338-373/Domain: EGF homology <EGF>
F:65/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #st

Query Match	5.9%;	Score 104;	DB 2;	Length 1251;
Best Local Similarity	24.3%;	Pred. No. 3.1;		
Matches 76;	Conservative 23;	Mismatches 100;	Indels 114;	Gaps

30 GGESIC SARALAKYITFTGKWSQTAPFKQYPLFRPPAQHSSLLGAHSSDYSMWRKNQY 89
 100 GGO--CSSRNCGLCPDETGRFCOV-----PAAGT---GACTGSSGCGCPDPRAM 143
 dbb

90 VSNGLRDFAERGEAWLMKEIEAA-----GEALQSVHAVFSAPVPSGTGQTSACL 140

Db 144 STGPLPLAPEGESVASKHAIYAVQVIADPPGEGPPQAHAAP---LVPLGPQOISA-- 198
 QY 141 EVQRHSLVSVFVRLVSPDFVGVDSLDLDCDGRWREQAALDLYPYDAGTDSFTSSP 200
 Db 199 EVQAPPPVNV--VRVHHPPPEASVQVHRIE-----GP 227
 QY 201 NFATIPQDTVTITSSSPSHPSFANSFYPRLKALPPIARVTLVLRQSP--RAFIAPPAPVL 258
 Db 228 N-AEGPASSQHLPLPKPOHPR-----PP-----TQPLGRCFQDTLPKQ 266
 QY 259 PSRNEIYDVSASVPETPL-----DCEVSLWSWGLCGGH-CGRLGTKSTRVVRVO-- 308
 Db 267 PCGSN-----PLPGLTKQEDCGSIGTAWGSKCHKCPQL-----QYTGVPK 309
 QY 309 -PANN--GSPCPE 318
 Db 310 VVVRGEVADCPQ 322

RESULT 11

minor coat protein - phase Pf3
 C:Species: phase Pf3
 A:Note: host Pseudomonas aeruginosa
 C:Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 23-Jul-1999
 C:Accession: A04232
 R:Luiten, R.G.M.; Buttermann, D.G.; Schoenmakers, J.G.G.; Konings, R.N.H.; Day, L.A.
 J. Virol. 56, 268-276, 1985
 A:Title: Nucleotide sequence of the genome of Pf3, an IncP-1 plasmid-specific filamentous phage.
 A:Reference number: A94693; MUID:85293231
 A:Accession: A04232
 A:Molecule type: DNA
 A:Residues: 1-483 <LUI>
 A:Cross-references: GB:M11912; NID:g215371; PIDN:AAA88379.1; PID:g215374
 C:Comment: Bacteriophage Pf3 is a class II filamentous phage.
 C:Comment: The host is strain O harboring IncP1 plasmids.
 C:Genetics:
 A:Gene: 483
 C:Superfamily: class II filamentous phage minor coat protein
 C:Keywords: capsid protein

Query Match 5.8%; Score 102; DB 1; Length 483;
 Best Local Similarity 25.8%; Pred. No. 1.4;
 Matches 62; Conservative 24; Mismatches 84; Indels 70; Gaps 14;

7 AAALGKALCALLAT-----LGAAGQPL--GGESICARALAKYSITFTGKWSQTAFPK 58
 112 ASGIGLGLAALLEAAADWVFDEGEIVKPLPGGGSPVLMRPVILNEYTVTGSAGQWSISK 171
 QY 59 QY-PLFRPPAOWSSLLG-----AAHSSDYSMWR-----KNQYVSNGLRDFAEGR- 101
 Db 172 ETEPDPRSPVGVYNGVNGPVVWSAVEDVGF--WRWYFADVLMDCGGRPNLVAYSDSGP 230
 QY 102 -EAWALMKEIEAAGEALQSVHAFVAPVSGTGTSAELEVQRHSLVSVFVVRVPSD 160
 Db 231 NEWQ-----DVGGYSLDSL-----PTEPEFVPLTDAELE----- 260
 QY 161 WFGVDSLDLDCDGRWREQAALDLYPYDAGTDSGFTFSFPNFIPTQDTVTITSSSPSH 220
 Db 261 --AGIDQYEPDPDWMR-----NLFFY-IEPDS-FTIETP-----IPSIDLSPVWSSSTNN 307

RESULT 12

D96566
 hypothetical protein F6D.21 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: D96566
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizaf, L.

Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: D96566
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-810 <STO>
 A:Cross-references: GB:AE005173; NID:g5903048; PIDN:AAU55607.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F6D8.21
 A:Map position: 1

Query Match 5.7%; Score 100.5; DB 2; Length 810;
 Best Local Similarity 25.7%; Pred. No. 3.5;
 Matches 47; Conservative 17; Mismatches 40; Indels 79; Gaps 10;

QY 129 VPSTGOTSAELEVRHRSILVSVFVVRIVPSDFVGVDSLDLDCDGRWREQAALDLYPYD 188
 Db 342 MPGG-----SRRRIVSF-----VGLDLDCG-----RYD 367
 QY 189 AGTDSGF-----TFSSPNFATIPQDTVTITSSSPSHPSFANSFYPRLKALPPIA-- 237
 Db 368 TPFSLSLFTLTAHDDFHQPNE-----TGAATKGGPREPWHDIHC-RLEG--PIAND 418
 QY 238 -----RVTLVRLRSRAFIPTPPAPVLSRNEI-----VD---SASVP 272
 Db 419 VLXNFQWRSRQGGKDLVKMRELGDIIIPSPVLFSEHDVWVNVQLFRSIDGAAAGFP 478
 QY 273 ETP 275
 Db 479 DSP 481

RESULT 13

T33152
 hypothetical protein T04D1.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33152
 R:Davidson, S.; Wohldmann, P.
 submitted to the EMBL Data Library, May 1998
 A:Description: The sequence of C. elegans cosmid T04D1.
 A:Reference number: 221292
 A:Accession: T33152
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-2957 <DAV>
 A:Cross-references: EMBL:AF067617; PIDN:AAAC17559.1; GSPDB:GN00019; CESP:T04D1.4
 A:Experimental source: strain Bristol N2; clone T04D1
 C:Genetics:
 A:Gene: CESP:T04D1.4
 A:Map position: 1
 A:Introns: 122/3; 293/3; 515/3; 1205/2; 1577/1; 2221/3; 2776/1; 2864/3

Query Match 5.7%; Score 100.5; DB 2; Length 2957;
 Best Local Similarity 18.8%; Pred. No. 17;
 Matches 63; Conservative 55; Mismatches 154; Indels 63; Gaps 10;

QY 7 AAALGKALCALLATLGAAGQPLGGES-----ICARALAKYSITFTGKWSQTAFPKQ 59
 Db 2407 AVALLAQVAAASASTSADQAQVAAQVQVLEALLAMSNPAALALGTSSSTGTSTPSA 2466
 QY 60 YPLFRPPAOWSSLLG-----AHSSDYSMWRKNQYVSNGLRDFAEGR-----EAWA 105
 Db 2467 PKSKPAPQKESTSSATATAAASAAQAAQAAQAAQAAQAAQAAQAAQAAQAAQAAQAA 2526

QY 106 LMKEIEAAGBALQ---SVHAFVAPAVPGSGTQTSAELEVVORRHSLVSVFVRIVSPDMF 162
Db 2527 IRLLELAGVGMQELAKMTMSDARVPMIHKOTRESLPVSKRPPIRLDTLTFVMSNPSWS 2586
QY 163 VGVDSLDCDGRWRQQAALDLYPDAGTSCFTFFSPNFATIPQDVTI--EITSSPSH 220
Db 2587 IDPLNEMSGN-----GTGSGAAPSCKAASRRPTPAATPLAVATPKPST 2631
QY 221 P--ANSFYPRKALPPIARVTLVRLRQSPRAFIPAPVPLPSRDNEIVDSASVPETPLDC 278
Db 2632 PKVEKSSATPSAPETFSASATCNKIGESLQ-FDDMISVFNKGTGELLAASKWPKA---T 2687
QY 279 EVSLW-----SSWGLCGGCGRLG 297
Db 2688 ELSAWLDANPDYNVHANSALFAHLSLGGAHSRIG 2722

RESULT 14
C70651
Species: Mycobacterium tuberculosis (strain H37Rv)
Accession: C70651
Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70651
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987
A: Accession: C70651
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-424 <COL>
A: Cross-references: GB: Z83866; GB: AL123456; NID: g3261691; PID: CAB06249.1; PID: e290921;
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: RV3074

Query Match 5.7%; Score 99.5; DB 2; Length 424;
Best Local Similarity 23.4%; Pred. No. 1.9; Mismatches 136; Indels 139; Gaps 22;
Matches 93; Conservative 29; Mismatches 136; Indels 139; Gaps 22;

QY 1 MENPSPAAALGKALCALL-----ATLGAAGQPLG--GESICSAKALAKYSITETGKWSQ 53
Db 24 LERLSAAAGQAAAAAAYDAAARAAAGAGVPAARRRGLASEIALARRD----- 74
QY 54 TAPKQYPLFRPPAQWSSILGAAHSSDYSM-----WRKNQYV-SNGLRDEFA 98
Db 75 -----SPARGSRHLGFAKALVEMPHITLAALDCGALSEMRAFLIIVRESACLDVA 123
QY 99 ER-----GEAWALMKEIAAGEALQSVHAFVSAPAV-----PSGTGQTSAELEVORRHSLV 149
Db 124 DRRALDAELCGDPPGLEMGDA-----RVVAARAATAYRLDPAQVVDRAANAENDR----- 174
QY 150 SFVRIVSPDMFVGVDSL-----DLCDGDRWRQQAALD-LYPYDAG 190
Db 175 --TVTIRPAPDTMTYLTALLPVAQGSVYAALTRAADTRCDG-RSRGQVMADTLVERVTG 231
QY 191 TDSGTFSSPNEATIPQDVTVEITSS-----SPSHPANSFY----PRKALPPIA----- 237
Db 232 RD-----AAVPTPIAVNLVMSDETLIGAANTPAQLCGYGPPIPAAVARTWVASAVT 281
QY 238 ----RVLVLRQSP-----RAFIPP---APVLPSSRDNEIVDSASVPETPLDCEV 280
Db 282 DQSRATLRLLYAHYPQAGALYMSERARLFPRLGLAAFIELDRQC-----RTIPY-CD 333
QY 281 SL-----WSSWGLCGGCGRLGTSRTRYRVQP 309
Db 334 PIRHRDHAHPWADGGTSAHNG-LGTCERCNYAKQAP 369

RESULT 15
146687
complement component C8 beta subunit - rabbit
C: Species: Oryctolagus cuniculus (domestic rabbit)
C: Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 20-Aug-1999
C: Accession: 146687
R: White, R.V.; Kaufman, K.M.; Letson, C.S.; Platteborze, P.L.; Sodetz, J.M.
J. Immunol. 152, 2501-2508, 1994
A: Title: Characterization of rabbit complement component C8: Functional evidence for
A: Reference number: 146686; MUID: 94179833
A: Accession: 146687
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 1-590 <WHI>
A: Cross-references: GB: L26980; NID: g469062; PID: AAA31192.1; PID: g469063
C: Superfamily: complement C9; EGF homology; LDL receptor ligand-binding repeat homolo
F: 63-117/Domain: thrombospondin type 1 repeat homology <THR>
F: 122-155/Domain: LDL receptor ligand-binding repeat homology <LDL>
F: 503-534/Domain: EGF homology <EGF>

Query Match 5.7%; Score 99.5; DB 2; Length 590;
Best Local Similarity 33.3%; Pred. No. 2.9;
Matches 28; Conservative 15; Mismatches 28; Indels 13; Gaps 6;

QY 246 QSPRAFIPPA---PVLPSRDNEIVDSASVPETPLDCEVSLWSWGLCGGCGRLGTSRT 302
Db 33 ERPSLSLEPTVVNRSLSAKRSRSDATPM---PIDCELSWSWWTMC-DPC-----QKKRY 84
QY 303 RYV-RVQPAN-NGSPCPPELEEEAE 324
Db 85 RHAYLLRPSQFNGEPCNFSKVE 108

Search completed: December 10, 2001, 20:03:09
Job time: 2731 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2001, 19:46:05 ; Search time 62.4 Seconds
(without alignments)
194.488 Million cell updates/sec

Title: US-09-170-042a-2

Perfect score: 1756

Sequence: 1 MENPSPAAALGKALCALLA.....NGSPCPLEEEAECPDNCV 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

al number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

*Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	462.5	26.3	807 1 FSP0_RAT	P35446 rattus norv
2	448.5	25.5	803 1 FSP0_XENLA	P35447 xenopus lae
3	127.5	7.3	741 1 NFL1_MOUSE	Q61985 mus musculus
4	117	6.7	772 1 NFL1_HUMAN	Q14494 h nuclear f
5	107	6.1	808 1 PLD_RICCO	Q41142 ricinus com
6	102	5.8	483 1 COA2_BPF3	P03624 bacterioph
7	100.5	5.7	810 1 PLD2_ARATH	Q95899 arabidopsis
8	99.5	5.7	590 1 COB8_RABIT	P98137 oryctolagus
9	99.5	5.7	812 1 PLD2_MAIZE	Q43270 zea mays (m
10	98.5	5.6	812 1 PLD2_BRAOC	P55939 brassica ol
11	97.5	5.6	808 1 PLD2_TOBAC	P93400 nicotiana t
12	97	5.5	1584 1 BAI1_HUMAN	O14514 homo sapien
13	96.5	5.5	598 1 BGAL_XANMN	P48982 xanthomonas
14	96.5	5.5	810 1 PLD1_BRAOC	O82549 brassica ol
15	96.5	5.5	812 1 PLD1_ORYSA	Q43007 oryza sativ
16	96	5.5	591 1 COB8_HUMAN	P07358 homo sapien
17	94	5.4	809 1 PLD1_VIGUN	O04865 vigna ungui
18	93.5	5.3	810 1 PLD1_ARATH	Q38882 arabidopsis
19	93	5.3	3027 1 POLG_PVEVL	Q05057 parsnip yel
20	92.5	5.3	912 1 PCGB_BOVIN	Q28062 bos taurus
21	92	5.2	1306 1 MSB2_YEAST	P32334 saccharomyc
22	91.5	5.2	341 1 FUN1_CRYFL	Q04469 cryptococcu
23	90	5.1	2265 1 FINE_BOVIN	P07589 bos taurus
24	89	5.1	1257 1 PER2_MOUSE	O54943 mus musculus
25	89	5.1	1567 1 FMN2_MOUSE	Q91104 mus musculus
26	88	5.0	867 1 SSPO_BOVIN	P98167 bos taurus
27	87.5	5.0	897 1 CHIA_ECOLI	P13656 escherichia
28	87	5.0	283 1 ISPE_CHILMU	Q961c0 chlamydia m
29	87	5.0	541 1 BBRI_SCHCO	P78741 schizophyll
30	87	5.0	863 1 AMPN_CAUCR	P37893 caulobacter
31	86.5	4.9	400 1 TX22_HUMAN	Q94458 homo sapien
32	86.5	4.9	590 1 INR1_MOUSE	P33896 mus musculus
33	86	4.9	940 1 MAZ4_SCHCO	P37938 schizophyll

RESULT 1

TSPO_RAT	FSP0_RAT	STANDARD;	PRT;	807 AA.
AC	P35446;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	F-SPONDIN PRECURSOR.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUP-Embryonic floor plate;			
RA	MEDLINE=92208952; PubMed=1555244;			
KL	Klar A., Baldassare M., Jessell T.M.;			
FT	"F-spondin: a gene expressed at high levels in the floor plate			
RT	encodes a secreted protein that promotes neural cell adhesion and			
RT	neurite extension.";			
RL	Cell 69:95-110(1992).			
CC	-!- FUNCTION: PROMOTES THE ATTACHMENT OF SPINAL CORD AND SENSORY			
CC	NEURON CELLS AND THE OUTGROWTH OF NEURITES IN VITRO. MAY			
CC	CONTRIBUTE TO THE GROWTH AND GUIDANCE OF AXONS IN BOTH THE SPINAL			
CC	CORD AND THE PNS.			
CC	-!- SUBCELLULAR LOCATION: SECRETED.			
CC	-!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE FLOOR PLATE.			
CC	-!- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcement/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; M88469; AAA1174.1; -			
DR	PIR; A38152; A38152.			
DR	InterPro; IPR002861; Reeler.			
DR	InterPro; IPR000884; TSP1.			
DR	Pfam; PF02014; Reeler; 1.			
DR	Pfam; PF00090; Tsp1; 6.			
DR	SMART; SM00209; TSP1; 6.			
DR	PROSITE; PS50092; TSP1; 5.			
DR	Glycoprotein; Signal; Repeat; Cell adhesion.			
FT	SIGNAL 1 28			POTENTIAL.
FT	CHAIN 29 807			F-SPONDIN.
FT	DOMAIN 443 494			TSP TYPE-1 1.
FT	DOMAIN 502 554			TSP TYPE-1 2.
FT	DOMAIN 559 610			TSP TYPE-1 3.
FT	DOMAIN 615 665			TSP TYPE-1 4.
FT	DOMAIN 669 720			TSP TYPE-1 5.
FT	DOMAIN 755 807			TSP TYPE-1 6.
FT	CARBOHYD 214 214			N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 681 681			N-LINKED (GLCNAC. . .) (POTENTIAL).

O60241 homo sapien
Q60519 mus musculus
P96920 mycobacteri
P07949 homo sapien
Q00610 homo sapien
P49951 bos taurus
P11442 rattus norv
Q28343 canis famli
Q00993 mus musculus
Q51389 pseudomonas
P42222 saccharomyc
Q04637 homo sapien

ALIGNMENTS

SQ SEQUENCE 807 AA; 90773 MW; 309525F9EAFEAE89A CRC64;
 Query Match 26.3%; Score 462.5; DB 1; Length 807;
 Best Local Similarity 34.2%; Pred. No. 4e-30;
 Matches 116; Conservative 49; Mismatches 137; Indels 37; Gaps 12;
 QY 9 ALGKALCALLATLGAAGOPLGESICSAALAKYISITFTGKWSQTAFFKQYPLFRPPAQ 68
 Db 176 SLTKKLCEQDPTLDGVTDRP---LDCCACGCTAKYRLTFYGNWSEKTHPKDYP--RRANH 230
 QY 69 WSELGAAHSSDYSMWRKNQVYNSGLRDFARGEAWALMKEIEAAG-EALQSVHAFV 127
 Db 231 WSAIIGSHSKNVLWEYGGYSEGKVAELGSPVKMEIEIRQOSDEVLTVIKAKAQP 290
 QY 128 -----AVPSGTGTSAELEVORHSLVSVFVVRTPSPDFVGVDSLDLDCGD-RWRE 178
 Db 291 SWQPVNVRAAP-----SAFSDVTRHLSFLTMGSPSPDNWGLSADLCTKCGWVQ 344
 QY 179 QAALDLYPYDAGTSGFTFSSPFATIPQDTVTEITSSSPSPANSFYPRKLPPIAR 238
 345 KVVQDLIPWDAGTSGVTYESPNKPTIPQEKIRPLTSL--DRPQSPFYDPEGSSITQVAR 402
 QY 239 VTLVRL-ROSPRAFIPAPVLSRNEIVDSASVPE-----TPLDCEVLSWSSWGLCGG 291
 Db 403 VVIERIARAGEQCNIVPDNV----DDIVADLA--PEEKDEDDTPTETCIYSNWPWSACGS 456
 QY 292 HGRCLCTKTRVVRVROPANNGSPCEPELEAEACVDPNC 330
 Db 457 STCEKGRMRQRLKQAQ-LDLSVPCPDTPQDFPCMGPGC 494
 RESULT 2
 FSPO_XENLA STANDARD; PRT; 803 AA.
 AC P35447.
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE F-SPONDIN PRECURSOR.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 MEDLINE=93376785; PubMed=8367492;
 Ruiz i Altaba A., Cox C., Jessell T.M., Klar A.;
 "Ectopic neural expression of a floor plate marker in frog embryos
 injected with the midline transcription factor Pintallavis";
 Proc. Natl. Acad. Sci. U.S.A. 90:8268-8272(1993).
 CC -1- FUNCTION: PROMOTES THE ATTACHMENT OF SPINAL CORD AND SENSORY
 CC NEURON CELLS AND THE OUTGROWTH OF NEURITES IN VITRO. MAY
 CC CONTRIBUTE TO THE GROWTH AND GUIDANCE OF AXONS IN BOTH THE SPINAL
 CC CORD AND THE PNS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE FLOOR PLATE.
 CC -1- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L09123; AAA19105.1; -
 DR PIR; A47723; A47723.
 DR InterPro; IPR002861; Reeler.
 DR InterPro; IPR000884; TSP1.

DR Pfam; PF02014; Reeler; 1.
 DR Pfam; PF00090; tsp.1; 6.
 DR SMART; SM00209; TSP1; 6.
 DR PROSITE; PS00092; TSP1; 6.
 KW Glycoprotein; Signal; Repeat; Cell adhesion.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 803 F-SPONDIN.
 FT DOMAIN 437 488 TSP TYPE-1 1.
 FT DOMAIN 496 548 TSP TYPE-1 2.
 FT DOMAIN 553 604 TSP TYPE-1 3.
 FT DOMAIN 609 661 TSP TYPE-1 4.
 FT DOMAIN 665 716 TSP TYPE-1 5.
 FT DOMAIN 751 803 TSP TYPE-1 6.
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 803 AA; 90702 MW; D3A54E329548AED9 CRC64;
 Query Match 25.5%; Score 448.5; DB 1; Length 803;
 Best Local Similarity 34.3%; Pred. No. 5.4e-29;
 Matches 115; Conservative 48; Mismatches 141; Indels 31; Gaps 12;
 QY 9 ALGKALCALLATLGAAGQPLGGESI---CSARALAKYISITFTGKWSQTAFFKQYPLFRP 65
 Db 172 SLTKRMCELDLTLEG-----GNEKTIPDCCACGTAKYRLTFYGNWSEKHAHPKDYP--RR 223
 QY 66 PAOWSSLLGAHSSDYSMWRKNQVYNSGLRDFARGEAWALMKEIEAAG-EALQSVHAFV 124
 Db 224 ANHWSAIIIGSHSGEYVLWEYGO--ASGVKQVAELGSPVKMEIEIRQKQDEVILTVIKAKA 282
 QY 125 SAPAVP--SGTGQTSAELEVORHSLVSVFVVRTPSPDFVGVDSLDLDCGD-RWREOAA 181
 Db 283 QWPAWQPLNVRAAPSAEFSVDRSRHLSFLAMGSPDNWGLTSEDLCCTKCGWQKV 342
 QY 182 LDLYPYDAGTSGFTFSSPFATIPQDTVTEITS--SPSPHANSFYPRKLPPIARV 240
 Db 343 QDLIPWDAGTSGVTYESPNKPTIPQEKIRPLTSLDHPQSPSMT----RGGPITPIARV 398
 QY 241 LVRLRQSPRAFIPAPVLPSPDNV-----DSASVPETPLDCEVLSWSSWGLCGGCGR 295
 Db 399 IERLARGE---QCNIIPDNVDIVADLVTEKDEDDTPTETCIYSNWPWSACSSATCD 454
 QY 296 LGTKSRTRYVRVOPANNGSPCEPELEAEACVDPNC 330
 Db 455 KGKRMQRMLKQAQ-LDLSVPCPDTPQDFPCMGPGC 488
 RESULT 3
 NFLI_MOUSE
 ID NFLI_MOUSE STANDARD; PRT; 741 AA.
 AC Q61985; O70234;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE NUCLEAR FACTOR ERYTHROID 2 RELATED FACTOR 1 (NF-E2 RELATED FACTOR 1)
 DE (NF-E2-RELATED FACTOR 1) (NUCLEAR FACTOR, ERYTHROID DERIVED 2, LIKE 1).
 GN NFE2L1 OR NFE1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 MEDLINE=95278942; PubMed=7759107;
 RA McKie J., Johnstone K., Mattei M.-G., Scambler P.;
 "Cloning and mapping of murine Nfe2l1";
 RL Genomics 25:716-719(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RX MEDLINE=98248571; PubMed=9580677;
 RA Prieschl E.E., Novotny V., Csonga R., Jaksche D., Elbe-Buerger A.,
 Thumb W., Auer M., Stingl G., Baumruker T.;

RT "A novel splice variant of the transcription factor Nrf1 interacts
 RL with the TNFalpha promoter and stimulates transcription.";
 CC Nucleic Acids Res. 26:2291-2297(1998).
 CC -1- FUNCTION: THE TNF ALPHA PROMOTER INTERACTS WITH THE EXTENDED KAPPA 3
 CC SITE OF THE TNF ALPHA PROMOTER AFTER FC GAMMA RIII STIMULATION AND
 CC PARTICIPATES IN THE INDUCTION OF THIS CYTOKINE. THE LONG ISOFORM
 CC IS EITHER INACTIVE OR REPRESSSES THE TRANSCRIPTIONAL ACTIVATION.
 CC -1- SUBUNIT: HETERODIMER (PROBABLE). IT MAY FORM A HETERODIMER WITH
 CC SEVERAL TRANSCRIPTION FACTORS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG (SHOWN HERE) AND A
 CC SHORT; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.
 CC -1- PTM: COULD BE DEPENDENT ON CKII PHOSPHORYLATION FOR BINDING.
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. CNC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X78709; CAA55362.1; -;
 CC EMBL: AF015881; AAC40108.1; -;
 CC HSSP: P34707; 1SKN.
 CC MGD: MGI:99421; Nfe2l1.
 CC InterPro: IPR001871; bZIP.
 CC Pfam: PF00170; bZIP; 1.
 CC SMART: SM00338; BRLZ; 1.
 CC PROSITE: PS00036; BZIP_BASIC; 1.
 CC DNA-binding; Nuclear protein; Phosphorylation; Alternative splicing.
 CC FT DOMAIN 125 259 ASP/GLU-RICH (ACIDIC).
 CC FT DOMAIN 414 447 POLY-SER.
 CC FT DOMAIN 467 486 BASIC MOTIF (BY SIMILARITY).
 CC FT DNA_BIND 628 643 LEUCINE-ZIPPER (BY SIMILARITY).
 CC FT DOMAIN 651 673 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC FT VARSPLIC 1 291 MISSING (IN SHORT ISOFORM).
 CC FT VARSPLIC 447 583 MISSING (IN SHORT ISOFORM).
 CC FT CONFLICT 318 318 S -> T (IN REF. 2).
 CC FT CONFLICT 387 387 P -> L (IN REF. 2).
 CC SEQUENCE 741 AA; 81545 MW; C01E89DD26E7CDEF CRC64;
 SQ
 Query Match 7.3%; Score 127.5; DB 1; Length 741;
 Best Local Similarity 22.4%; Pred. No. 0.0073;
 Matches 79; Conservative 45; Mismatches 119; Indels 109; Gaps 17;
 QY 19 LATLGAAGQPLGGESICARALAKYSITFTG-KWSQTAFPKQYPLFRPPAQWSSLLGAAH 77
 Db 153 LEDLGAVAPPVSGD-----LTKEIDILDLILWRQDID-----LGAGR 189
 QY 78 S-SDYSWMRKNOYVSNGLRDFAEAGW-----ALMKEIEAAGEALQSVAHFSA--PA 128
 Db 190 EVDYSHRQKEQDVDELQDQGEREDTWSGEGAEALARDLLVDGTGSEFPFAOFFADVSS 249
 QY 129 VPSGTGQTSAELEVRHSLVSFVVRVPSDPWFGVDSLDICDGRWRE-QAALDLYPY 187
 Db 250 IPEAVPSESPALQ--NSLSPLLTGTESP-----FDL--EQQWQDLMSINEMQAM 297
 QY 188 DACTD-SGTFSSPNPATIPQDTVTETITSSSPSHAN-----SFYVPR 229
 Db 298 EVNTSASEILYNAP-----PGDPLSSNYS LAPNTPINQVNSLRQASLGCSQDFSLFSPE 352
 QY 230 LKALPPIARTVLRVLRQSPRAFIP-----PAPVLPSEDRNEIVDSASVPETP----- 275
 Db 353 VESLPVASSSTLLPVPFNSNSTLUNSTFGTNIAGFPFPFSQLNGTANDTSGPELPDLOGL 412
 QY 276 -----LDCEVSLWSSWGLCGGCHGRGLTGRTRRYRVQVQPNNGSPCPPELEEE 322
 Db 413 LDAMLD-EISLMD-----LAIEGFNPVQASQLEEE 443

RESULT 4
 NFE1L1_HUMAN STANDARD; PRT; 772 AA.
 AC Q1494; Q12877;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NUCLEAR FACTOR ERYTHROID 2 RELATED FACTOR 1 (NF-E2 RELATED FACTOR 1)
 DE (NF-E2-RELATED FACTOR 1) (NUCLEAR FACTOR, ERYTHROID DERIVED 2, LIKE 1)
 DE (TRANSCRIPTION FACTOR 1) (TRANSCRIPTION FACTOR HBZ17) (TRANSCRIPTION
 DE FACTOR LCB-F) (LOCUS CONTROL REGION-FACTOR 1).
 GN NFE2L1 OR NFE1 OR YCF11 OR HBZ17.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 DX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95095252; PubMed=8001966;
 RA Luna L., Johnson O., Skartlien A.H., Pedoutour F., Turc-Carel C.,
 RA Prydz H., Kollstoe A.-B.;
 RT "Molecular cloning of a putative novel human bZIP transcription
 RT factor on chromosome 17q22.";
 RL Genomics 22:553-562(1994).
 RN [2]
 RP SEQUENCE OF 326-772 FROM N.A.
 RX MEDLINE=94310069; PubMed=8036168;
 RA Caterina J.J., Donze D., Sun C.W., Ciavatta D.J., Townes T.M.;
 RT "Cloning and functional characterization of LCR-F1: a bZIP
 RT transcription factor that activates erythroid-specific, human globin
 RT gene expression.";
 RL Nucleic Acids Res. 22:2383-2391(1994).
 CC -1- FUNCTION: ACTIVATES ERYTHROID-SPECIFIC, GLOBIN GENE EXPRESSION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. CNC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X77366; CAA54555.1; -;
 CC EMBL: U08853; AAA20466.1; -;
 CC HSSP: P34707; 1SKN.
 CC MIM: 163260; -;
 CC InterPro: IPR001871; bZIP.
 CC Pfam: PF00170; bZIP; 1.
 CC SMART: SM00338; BRLZ; 1.
 CC PROSITE: PS00036; BZIP_BASIC; 1.
 CC Transcription regulation; Activator; DNA-binding; Nuclear protein.
 CC FT DOMAIN 125 288 ASP/GLU-RICH (ACIDIC).
 CC FT DOMAIN 496 517 POLY-SER.
 CC FT DNA_BIND 659 674 BASIC MOTIF.
 CC FT DOMAIN 682 704 LEUCINE-ZIPPER.
 CC SEQUENCE 772 AA; 84703 MW; C868807C6046BEF5 CRC64;
 Query Match 6.7%; Score 117; DB 1; Length 772;
 Best Local Similarity 23.3%; Pred. No. 0.055;
 Matches 78; Conservative 35; Mismatches 110; Indels 112; Gaps 17;
 QY 19 LATLGAAGQPLGGESICARALAKYSITFTG-KWSQTAFPKQYPLFRPPAQWSSLLGAAH 77
 Db 153 LEDLGAVAPPVSGD-----LTKEIDILDLILWRQDID-----LGAGR 189
 QY 78 S-SDYSWMRKNOYVSNGLRDFAEER-----GE-AWALMKEIEAAGEALQSVAHFSAVAPS 131
 Db 190 EVDYSHRQKEQDVDELQDQGEREDTWSGEGAEALARNLLVDGTGSEFPA-----QVPS 244

QY 132 GYGQTSAELEVORR-----HSLVSFVVRIVPSPD-----WFGV 165
| | | | | : : : : :
| | | | | : : : : :
Db 245 GEDQALSECLRLLEATCPFGENAEPPADISSITEAVPSEPPALQNNLLSPLLTGT 304
| | | | | : : : : :
| | | | | : : : : :
QY 166 DS-LDLCGDWRRE-OAALDLYPDAGTD-SGTFESSPNFATIPQDVTITSSPSHPA 222
| | | | | : : : : :
| | | | | : : : : :
Db 305 ESPFDL--EQWQDLMSIMEMQAEVNTSASEILYSAP-----PCDPLSTNYSIAPNTP 357
| | | | | : : : : :
| | | | | : : : : :
QY 223 NS-----EYVRLKALPPPTARTVTLVRLSPRA-----F 251
| | | | | : : : : :
| | | | | : : : : :
Db 358 NONVSLHOASLGCGSQDQLLSPVESLPVASSSTLLPLAPSNTSLNFTGSTNLTLGLF 417
| | | | | : : : : :
| | | | | : : : : :
QY 252 TPP-----APVLPSPRNEIVDSASVPETPL 276
| | | | | : : : : :
| | | | | : : : : :
Db 418 FPPQLNGTANDTAGPELPDPLGLLDEAMLDEISL 452
| | | | | : : : : :
| | | | | : : : : :
RESULT 5
PLD_RICCO STANDARD; PRT; 808 AA.
Q41142; P93507;
01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE PHOSPHOLIPASE D PRECURSOR (EC 3.1.4.4) (PLD) (CHOLINE PHOSPHATASE)
DE (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CV. HALE; TISSUE=Endosperm;
RX MEDLINE=94327597; PubMed=8051126;
RA Wang X., Xu L., Zheng L.;
RT "Cloning and expression of phosphatidylcholine-hydrolyzing
phospholipase D from Ricinus communis L.";
RL J. Biol. Chem. 269:20312-20317(1994).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=97134969; PubMed=9880529;
RA Xu L., Zheng L., Coughlan S.J., Wang X.;
RT "Structure and analysis of phospholipase D gene from Ricinus communis
L.";
RL Plant Mol. Biol. 32:767-771(1996).
RN [3]
FUNCTION: PLAYS AN IMPORTANT ROLE IN CELLULAR PATHWAYS INCLUDING
SIGNAL TRANSDUCTION PATHWAYS.
CC CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
PHOSPHATIDATE.
CC COFACTOR: CALCIUM (BY SIMILARITY).
CC TISSUE SPECIFICITY: MOSTLY EXPRESSED IN VACUOLES, ENDOPLASMIC
RETICULUM, A FEW IN PLASTIDS AND PLASMA MEMBRANE. EXPRESSION IS
HIGHER IN RADICLE THAN IN ENDOSPERM.
CC SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC EMBL; L33686; AAB04095.1; -;
DR EMBL; U72693; AAB37305.1; -;
DR InterPro; IPR000008; C2.
DR Pfam; PF00614; PLDc; 2.
DR SMART; SM00239; C2; 1.
DR SMART; SM00155; PLDc; 2.

DR PROSITE; PS50004; C2_DOMAIN_2; 1.
KW Hydrolase; Lipid degradation; Calcium.
FT PROPEP 1 30
| | | | | : : : : :
| | | | | : : : : :
FT CHAIN 31 808 PHOSPHOLIPASE D.
| | | | | : : : : :
| | | | | : : : : :
FT DOMAIN 1 109 C2 DOMAIN.
| | | | | : : : : :
| | | | | : : : : :
FT CONFLICT 268 288 L -> I (IN REF. 2).
SQ SEQUENCE 808 AA; 91992 MW; E75F6CFFB9ADF3CB CRC64;
Query Match 6.1%; Score 107; DB 1; Length 808;
Best Local Similarity 24.6%; Pred. No. 0.38;
Matches 59; Conservative 24; Mismatches 81; Indels 76; Gaps 14;
QY 120 VHAVFSAPVPGTGTQTSAELEVQ---RRHSLVSFVVRIVPSPD-----WFGVDSLD 169
| | | | | : : : : :
| | | | | : : : : :
Db 303 VHCVL-CPRPDDGGSEVQDLQISTMTTHQKIVVVDVSAMPNGDSQRRRIVSEVG--GLD 359
| | | | | : : : : :
| | | | | : : : : :
QY 170 LCGDRWRQAAALDLYPDAGTDSGF-----TFSSPNFATIPQDVTITSSSPSH 220
| | | | | : : : : :
| | | | | : : : : :
Db 360 LCDG-----RYDSPFHSFLFTLDSAHDDFHQPFAG-----ASIERGGPRE 401
| | | | | : : : : :
| | | | | : : : : :
QY 221 PANSFYVRLKALPPPTA-----RVTLVRLSPRAFPAPVLPSPRDN 264
| | | | | : : : : :
| | | | | : : : : :
Db 402 PWHDI-HSRLEG--PIANDVLFNFQWRKQGGKOLLQIOLRELDVILPPSPVMPDDFE 458
| | | | | : : : : :
| | | | | : : : : :
QY 265 I-----VDSASV---PETPLDCEVSWSSWGCGHCGRLGKTRTRYVR-VQPA 312
| | | | | : : : : :
| | | | | : : : : :
Db 459 AWWVQLFRSIDGGAARFPETPEDAPEA-----GLVSGKDNIDRSIQDAYIHAIIRAKN 513
| | | | | : : : : :
| | | | | : : : : :
RESULT 6
COA2_BPPF3 STANDARD; PRT; 483 AA.
ID COA2_BPPF3
AC P03624;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-1989 (Rel. 10, Last annotation update)
DE MINOR COAT PROTEIN (ORF 483).
OS Bacteriophage Pf3.
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
OX NCBI_TaxID=10872;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=NEW-YORK, AND NIJMEGEN;
RX MEDLINE=85293231; PubMed=3928901;
RA Luitert R.G.M., Putterman D.G., Schoenmakers J.G.G., Konings R.N.H.,
RA Day L.A.;
RT "Nucleotide sequence of the genome of Pf3, an incp-1 plasmid-specific
filamentous bacteriophage of Pseudomonas aeruginosa.";
RL J. Virol. 56:268-276(1985).
CC MISCELLANEOUS: THE STRAIN NIJMEGEN SEQUENCE IS SHOWN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC EMBL; M11912; AAA88379.1; -;
DR EMBL; M19377; AAA88388.1; -;
DR PIR; A04232; VCBPI3.
KW Coat protein.
FT VARIANT 419 419 L -> F (IN STRAIN NEW-YORK).
SQ SEQUENCE 483 AA; 52197 MW; DCBAA598BA700BB CRC64;

Query Match 5.8%; Score 102; DB 1; Length 483;
Best Local Similarity 25.8%; Pred. No. 0.52;
Matches 62; Conservative 24; Mismatches 84; Indels 70; Gaps 14;
QY 7 AALGKALCALLIAT-----LGAAGQPL--GGESICSARALAKYSITFTGWSQTAPFK 58

```

DB 112 ASGIGLGAALLAEDWDFDEGETVKGPGSPVLMRPVILNEXVTVTSGAGOWSTSK 171
QY 59 QY-PLFRPPAOWSSLLG-----AAHSSDYSNMR-----KNOYVSNGLRDFAEGR- 101
DB 172 EYEPDPRSPGWYNGNPNVWVSAVEDVGF-RRYWFADVLMDGGRPNVLYVAYSDSGP 230
QY 102 -BAWALMKEIEAAGRALQSVHAFVAPVPGTGTGTSAELEVORRHSLVSVFVRIVPSD 160
DB 231 NEYQW-----DVGGYSLDSL-----PTEPEFVPLTDAELE-----260
QY 161 WFGVDSLDLDCGDRWRQEAALDLYPDAGTDSGTFTSPNFATIPQDVTVTITSSSPSH 220
DB 261 --AGIDQYQYEPDPDWR-----NLFPY-IEPDS-FTIETP-----IPSLDLPVSSSSNN 307

RESULT 7
PLD2_ARATH STANDARD; PRT; 810 AA.
ID Q93SQ9;
20-AUG-2001 (Rel. 40, Created)
20-AUG-2001 (Rel. 40, Last sequence update)
20-AUG-2001 (Rel. 40, Last annotation update)
DE 2) (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D 2).
GN PLD2 OR ATIG5270 OR F6D8.21.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Alcafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vyotskaya V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN VARIOUS CELLULAR PROCESSES.
CC -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
CC PHOSPHATIDATE.
CC -!- COFACTOR: CALCIUM (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AC008016; RAD55607.1;
CC InterPro; IPR000008; C2.
CC DR InterPro; IPR001736; PLD.
CC DR Pfam; PF00614; PLDC; 2.
CC SMART; SM00239; C2; 1.

```

```

DR SMART; SM00155; PLDC; 2.
DR PROSITE; PS0004; C2_DOMAIN_2; FALSE_NEG.
KW Hydrolase; Lipid degradation; Calcium.
FT PROPEP 1 36 BY SIMILARITY.
FT CHAIN 37 810 PHOSPHOLIPASE D 2.
FT DOMAIN 1 110 C2_DOMAIN.
FT SEQUENCE 810 AA; 91597 MW; FLFA36D8C9CFBCCA CRC64;

Query Match 5.7%; Score 100.5; DB 1; Length 810;
Best Local Similarity 25.7%; Pred. No. 1.3;
Matches 47; Conservative 17; Mismatches 40; Indels 79; Gaps 10;

QY 129 VPSGTGTSAELEVORRHSLVSVFVRIVPSDFVGVDSLDLDCGDRWRQEAALDLYPD 188
DB 342 MPSGG-----SRSRIVSF-----VGGDLDCG-----RYD 367

QY 189 AGTDSGF-----TFSSPNFATIPQDVTVTITSSSPSHSPANSFYPRKALPPIA-- 237
DB 368 TPFHSLFRLDTAHHDDFHOFNF-----TGAALTGGPREPWHDIHC-LEG--PIAWD 418

QY 238 -----RVTLVRLQSPRAFPAPVLPSPRNEI-----VD---SASVP 272
DB 419 VLYNFEORWSROGKDIIVKMRGLDIIIPSPVLFSEHDVWVQLFRSIDGGAAGFP 478

QY 273 ETP 275
DB 479 DSP 481

RESULT 8
CO8B_RABIT STANDARD; PRT; 590 AA.
ID AC P98137;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COMPLEMENT COMPONENT C8 BETA CHAIN PRECURSOR.
GN C8B.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
RX MEDLINE=94179833; PubMed=7510745;
RX White R.V., Kaufman K.M., Letson C.S., Platteborze P.L.,
RX Sodetz J.M.;
RA "Characterization of rabbit complement component C8. Functional
RA evidence for the species-selective recognition of C8 alpha by
RA homologous restriction factor (CD59).";
RL J. Immunol. 152:2501-2508(1994).
CC -!- FUNCTION: C8 IS A CONSTITUENT OF THE MEMBRANE ATTACK COMPLEX.
CC C8 BINDS TO THE C5B-7 COMPLEX, FORMING THE C5B-8 COMPLEX. C5-B8
CC BINDS C9 AND ACTS AS A CATALYST IN THE POLYMERIZATION OF C9.
CC -!- SUBUNIT: C8 IS COMPOSED OF THREE CHAINS: ALPHA, BETA AND GAMMA.
CC THE BETA CHAIN BINDS TO THE C8 ALPHA CHAIN AND TO THE C5B-C7
CC COMPLEX, PRESUMABLY TO C5B. IT IS ESSENTIAL TO THE INCORPORATION
CC OF C8 INTO THE C5B-C8 COMPLEX.
CC -!- SIMILARITY: TO COMPLEMENT FACTORS C6, C7, C9, AND TO PERFORIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC

```

(Oryza sativa L.) and cloning of cDNA for PLD from rice and maize (Zea mays L.).
Plant Cell Physiol. 36:903-914 (1995).
- FUNCTION: PLAYS AN IMPORTANT ROLE IN VARIOUS CELLULAR PROCESSES.
- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A PHOSPHATIDATE.
- COFACTOR: CALCIUM (BY SIMILARITY).
- SUBUNIT: MONOMER (BY SIMILARITY).
- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
EMBL; D73410; BAAL1135.1; -
MaizeDB; L13853; -
InterPro: IPR000008; C2.
InterPro: IPR001736; PLD.
Pfam: PF00614; PLDC; 2.
SMART; SM00239; C2; 1.
SMART; SM00155; PLDC; 2.
PROSITE; PS50004; C2.DOMAIN.2; 1.
PROPEP 1 ?
CHAIN ? 812 PHOSPHOLIPASE D.
DOMAIN 1 114 C2 DOMAIN.
SEQUENCE 812 AA; 92242 MW; D05CB351655BCC61 CRC64;
Query Match 5.7%; Score 99.5; DB 1; Length 812;
Best Local Similarity 25.7%; Pred. No. 1.6;
Matches 49; Conservative 19; Mismatches 62; Indels 61; Gaps 12;
QY 127 PAVPSGTGOTSAREVQ---RRHSLVSEFVVRVPSPD-----WFGVDSIDLCDGDRW 176
Db 313 PRNPDDSGSFVQDLISTFTTHQKIVVDHMPNQSGQRRIVSFIG--GIDLCDG--RY 369
QY 177 REQ-----AALDLPPYDAGTSGFTFSSPNFATIPQDTVTETSSPSHPANSFYPRLK 231
Db 370 DTQVHSLFTLDTVHHD-----DFHQPNFEG-----GSIKKGGPREPHDI-HSLE 415
QY 232 ALDPPIA-----RVTLVRLRQSPRAFIPAPVLPDRNEI-----VD 267
Db 416 G--PIAWDVLNFEQRWRKGGKGLLRLDLPLDIIIPSPVMFPEDRETWNVQLFRSID 473
QY 268 SASV---PETP 275
Db 474 GGAAGFGPETP 484
RESULT 10
PLD_MAIZE ID PLD_MAIZE STANDARD; PRT; 812 AA.
AC Q43270; PLD2_BRAOC STANDARD; PRT; 812 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PHOSPHOLIPASE D PRECURSOR (EC 3.1.4.4) (PLD) (CHOLINE PHOSPHATASE)
DE (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D).
DE Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Panicoideae; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Zea mays (Maize).
ON NCBI_TaxID=4577;
RN [1]
RC STRAIN=CV, MO17;
RX MEDLINE=96012933; PubMed=7551587;
RA Ueki J., Morioka S., Komari T., Kumashiro T.;
RA "Purification and characterization of phospholipase D (PLD) from rice


```

RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-38.
RC STRAIN=7AS1.
RX MEDLINE=96121416; PubMed=8563148;
RA Taron C.H., Benner J.S., Hornstra L.J., Guthrie E.P.;
RT "A novel beta-galactosidase gene isolated from the bacterium
RT xanthomonas manihotis exhibits strong homology to several eukaryotic
RT beta-galactosidases."
RL Cytobiology 5:603-610(1995).
CC -!- FUNCTION: PREFERENTIALLY HYDROLYZES BETA(1->3) GALACTOSYL LINKAGES
CC OVER BETA(1->4) LINKAGES. HAS A PH OPTIMUM OF 4.5.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L35444; AAC41485.1; -
CC InterPro; IPR001944; Glyco_hydro_35.
CC Pfam; PF01301; Glyco_hydro_35; 1.
CC PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
CC KW Hydrolase; Glycosidase; Signal.
CC FT SIGNAL 1 21
CC FT CHAIN 22 598 BETA-GALACTOSIDASE.
CC FT ACT_SITE 184 184 PROTON DONOR (POTENTIAL).
CC FT ACT_SITE 260 260 NUCLEOPHILE (POTENTIAL).
CC SEQUENCE 598 AA; 66085 MW; DB4C3F05E01435BF CRC64;

Query Match 5.5%; Score 96.5; DB 1; Length 598;
Best Local Similarity 26.8%; Pred. No. 1.9;
Matches 38; Conservative 14; Mismatches 53; Indels 37; Gaps 7;

QY 161 WF-----VGVDLDCDGRW--REQAALDLYPDYDAGTSGFTFSSPNFATIPQDRTV 211
DB 265 WFDHNGKPHAATDARQQAEEFWILRQGSANLYMFIGTSGFG-MNCANFQNPDSHYA 323
QY 212 EITSS-----SPSHPANSPYPRKALPPIARVTIVLRQSPRAPIPPAPVLP SRDN 263
DB 324 PQTTSYVDYDAILDAGHPKPKALMKR---DAIARVTGVQ-----PPALPAP----- 366
QY 264 EIVDSASVPETPLDCEVLSWSS 285
DB 367 --ITTTLPATPLRESASLWDN 386

RESULT 14
PLD1_BRAOC
ID PLD1_BRAOC STANDARD; PRT; 810 AA.
AC 082549;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHOLIPASE D 1 PRECURSOR (EC 3.1.4.4) (PLD 1) (CHOLINE PHOSPHATASE
DE 1) (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D 1).
GN PLD1.
OS Brassica oleracea var. capitata (Cabbage).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3716;
RN [1]
RP SEQUENCE FROM N.A.
RA Pannenberg I., Mansfeld J., Ulbrich-Hofmann R.;
RT "Cloning of phospholipase D from Brassica oleracea var. capitata.";
RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN VARIOUS CELLULAR PROCESSES.
CC -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A

```

```

CC PHOSPHATIDATE.
CC -!- COFACTOR: CALCIUM (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF090445; AAC78487.1; -
CC EMBL; AF113918; AAD17208.1; -
CC InterPro; IPR000008; C2.
CC InterPro; IPR001736; PLD.
CC Pfam; PF00614; PLDc; 2.
CC SMART; SM00239; C2; 1.
CC SMART; SM00155; PLDC; 2.
CC PROSITE; PS50004; C2_DOMAIN_2; FALSE_NEG.
CC Hydrolase; Lipid degradation; Calcium.
CC PROPEP 1 36 BY SIMILARITY.
CC FT CHAIN 37 810 PHOSPHOLIPASE D 1.
CC FT DOMAIN 1 110 C2 DOMAIN.
CC SQ SEQUENCE 810 AA; 91837 MW; CBC5B3D7E1F0DAAB CRC64;

Query Match 5.5%; Score 96.5; DB 1; Length 810;
Best Local Similarity 25.1%; Pred. No. 2.7;
Matches 51; Conservative 19; Mismatches 62; Indels 71; Gaps 12;

QY 120 VHAFSAPAVPSGTGQTSAELEVQ---RRHSLVSFVVRIVPSPDW-----FVGVDSL 168
DB 304 VHCIL-CPRPDDGGSGIVQNLQYSAMETHQKIVVDSEMPSSRGSGMRIVSEVG--GI 360
QY 169 DLCDGRNRQOALDLYPYDAGTDSGF-----TFSSPNFATIPQDRTVITSSSPS 219
DB 361 DLCDG-----RYDTPFHSLSFRTLDTVHDDHQPNF-----TGAATKGGPR 402
QY 220 HPANSEFYPRKALPPIA-----RVTLVRLSPRAPIPPAPVLP SRDN 263
DB 403 EPWHDIT-HSRLEG--PIANDVLYNFQNRKSGQKDLVLRKLSLIIITPSVPMQEDH 459
QY 264 EI-----VD-----SASVPETP 275
DB 460 DVWNVQLFRSIDGGAAGFPESP 482

RESULT 15
PLD1_ORYSA
ID PLD1_ORYSA STANDARD; PRT; 812 AA.
AC 043007;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PHOSPHOLIPASE D 1 PRECURSOR (EC 3.1.4.4) (PLD 1) (CHOLINE PHOSPHATASE
DE 1) (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D 1).
GN PLD1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
AC STRAIN=CV. KOSHIHIKARI.
RX MEDLINE=96012933; PubMed=7551587;
RA Ueki J., Morioka S., Komari T., Kumashiro T.;
RT "Purification and characterization of phospholipase D (PLD) from rice
RT (Oryza sativa L.) and cloning of cDNA for PLD from rice and maize
RT (Zea mays L.).";
RN Plant Cell Physiol. 36:903-914(1995).

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2001, 19:14:28 ; Search time 101.35 Seconds
(without alignments)
477.712 Million cell updates/sec

Title: US-09-170-042a-2
Perfect score: 1756
Sequence: 1 MENSPAAALGKALCALLLA.....NGSPCELEEEACVDPNCV 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

1 number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1749	99.6	331	4 Q9ULW1	Q9ULW1 homo sapien
2	1741	99.1	331	4 Q9BUD6	Q9BUD6 homo sapien
3	1520.5	86.6	330	11 Q9WV75	Q9WV75 rattus norv
4	1108	63.1	331	13 Q42112	Q42112 brachydanio
5	868	49.4	334	13 Q42111	Q42111 brachydanio
6	528.5	30.1	601	5 Q9V746	Q9V746 drosophila
7	518	29.5	598	5 Q02029	Q02029 drosophila
8	470.5	26.8	808	13 Q42113	Q42113 brachydanio
9	464.5	26.5	802	13 Q9W770	Q9W770 gallus gall
10	460.5	26.2	807	4 Q9HCB6	Q9HCB6 homo sapien
11	460.5	26.2	807	6 Q9GLX9	Q9GLX9 bos taurus
12	457	26.0	624	4 Q94862	Q94862 homo sapien
13	457	26.0	898	5 Q76822	Q76822 branchiosto
14	445.5	25.4	763	5 Q9XZD0	Q9XZD0 drosophila
15	443	25.2	803	13 Q42114	Q42114 brachydanio
16	439.5	25.0	873	5 Q9V692	Q9V692 drosophila
17	422.5	24.1	1682	5 Q9V693	Q9V693 drosophila
18	420.5	23.9	216	4 Q9H711	Q9H711 homo sapien
19	406	23.1	805	5 Q19305	Q19305 caenorhabdi

20	144.5	8.2	1536	4 Q9C014	Q9C014 homo sapien
21	135	7.7	446	4 Q43384	Q43384 homo sapien
22	134	7.6	1290	4 Q9UP26	Q9UP26 homo sapien
23	128	7.3	687	5 Q23729	Q23729 cryptospori
24	125	7.1	660	5 Q23832	Q23832 cryptospori
25	123.5	7.0	749	11 Q922E8	Q922E8 mus musculu
26	121.5	6.9	238	5 Q76510	Q76510 cryptospori
27	117.5	6.7	770	5 Q20942	Q20942 caenorhabdi
28	114	6.5	1423	5 Q9WIA0	Q9WIA0 drosophila
29	112	6.4	543	4 Q60407	Q60407 homo sapien
30	109	6.2	168	5 Q9G222	Q9G222 cryptospori
31	108	6.2	123	5 Q27550	Q27550 cryptospori
32	108	6.2	123	5 Q9TVV7	Q9TVV7 cryptospori
33	108	6.2	765	5 Q908J9	Q908J9 neospora ca
34	107.5	6.1	1523	3 Q9HFX4	Q9HFX4 candida alb
35	106	6.0	168	5 Q9G223	Q9G223 cryptospori
36	105	6.0	123	5 Q76782	Q76782 cryptospori
37	105	6.0	123	5 Q9TZP9	Q9TZP9 cryptospori
38	104.5	6.0	440	5 Q18003	Q18003 caenorhabdi
39	103	5.9	265	11 Q9Z132	Q9Z132 mus musculu
40	103	5.9	1253	11 Q61810	Q61810 mus musculu
41	102.5	5.8	588	13 Q9PVW7	Q9PVW7 paralicthy
42	102.5	5.8	738	5 Q9W321	Q9W321 drosophila
43	102.5	5.8	809	10 Q9SDZ6	Q9SDZ6 lycopersico
44	102	5.8	168	5 Q9G221	Q9G221 cryptospori
45	102	5.8	1863	11 Q9JLQ1	Q9JLQ1 mus musculu

ALIGNMENTS

RESULT 1

Q9ULW1 ID Q9ULW1 PRELIMINARY; PRT; 331 AA.
AC Q9ULW1;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SPONDIN 2.
GN SPON2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99443867; PubMed=10512675;
RA Manda R., Kohno T., Matsuno Y., Takenoshita S., Kuwano H., Yokota J.;
RT "Identification of genes (SPON2 and C20orf2) differentially expressed
RT between cancerous and noncancerous lung cells by mRNA differential
RT display.";
RL Genomics 61:5-14(1999).
DR EMBL; AB027466; BAA85892.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF000090; tsp_1; 1.
DR PROSITE; PS50092; TSP1; 1.
DR SMART; SM00209; TSP1; 1.
SQ SEQUENCE 331 AA; 35772 MW; FLD6F0BAC1412CA7 CRC64;

Query Match 99.6%; Score 1749; DB 4; Length 331;
Best Local Similarity 99.7%; Pred. No. 6.4e-140;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MENSPAAALGKALCALLATLGAAGQPLGSGESCSARALAKYSITFTGKWSATAPKQY 60
|||||
Qb 1 MENSPAAALGKALCALLATLGAAGQPLGSGESCSARAPAKYSITFTGKWSATAPKQY 60
|||||
QY 61 PLFRPPAQWSSLLGAHSSYSMMRKQYNSGLRDPFAERGEAWALMKEETAAGEALQSV 120
|||||
Qb 61 PLFRPPAQWSSLLGAHSSYSMMRKQYNSGLRDPFAERGEAWALMKEETAAGEALQSV 120
|||||
QY 121 HAVFSAFAVSGTQCTSAELEVRHSLVSVFVVRIVSPDFVGVDSLDLDCDGRWREQA 180

Db 121 HAVFSAPAVPGTGTSAELEVQRHSLVSFVVRIVSPDWFVGVDSLDLDCGDRWREQA 180
Qy 181 ALDLYPYDAGTDSGFTSSPNFATIPQDVTETITSSPSHPANSFYYPRLKALPPIARVT 240
Db 181 ALDLYPYDAGTDSGFTSSPNFATIPQDVTETITSSPSHPANSFYYPRLKALPPIARVT 240
Qy 241 LVRLRQSPRAFIPAPVLPSPRDNEIVDSASVPETPLDCEVSLWSSWGLCGHCGRLGTSK 300
Db 241 LVRLRQSPRAFIPAPVLPSPRDNEIVDSASVPETPLDCEVSLWSSWGLCGHCGRLGTSK 300
Qy 301 RTRVYRVQPNANGSPCPPELEEEACVPDNCV 331
Db 301 RTRVYRVQPNANGSPCPPELEEEACVPDNCV 331

RESULT 2

Q9BUD6 PRELIMINARY; PRT; 331 AA.
AC Q9BUD6;
ID Q9BUD6;
Qy 01-JUN-2001 (Tremblrel. 17, Created)
Db 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE SPONDIN 2, EXTRACELLULAR MATRIX PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-CHORIOCARCINOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002707; AA02707.1;
KW Matrix protein.
SQ SEQUENCE 331 AA; 35844 MW; 418E244B893C59F4 CRC64;

Query Match 99.1%; Score 1741; DB 4; Length 331;
Best Local Similarity 99.1%; Pred. No. 3e-139;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MENPSAALGKALLATLGAAGPLGSESTCSARALAKYSITITGKWSQTAPFKQY 60
Db 1 MENPSAALGKALLATLGAAGPLGSESTCSARAPAKYSITITGKWSQTAPFKQY 60
Qy 61 PLFRPPAOWSSLLGAHSSDYSMRKNQYVSNGLRDFAEERGEAWALMKEIEAAGEALQSV 120
Db 61 PLFRPPAOWSSLLGAHSSDYSMRKNQYVSNGLRDFAEERGEAWALMKEIEAAGEALQSV 120
Qy 121 HAVFSAPAVPGTGTSAELEVQRHSLVSFVVRIVSPDWFVGVDSLDLDCGDRWREQA 180
Db 121 HEVFSAPAVPGTGTSAELEVQRHSLVSFVVRIVSPDWFVGVDSLDLDCGDRWREQA 180
Qy 181 ALDLYPYDAGTDSGFTSSPNFATIPQDVTETITSSPSHPANSFYYPRLKALPPIARVT 240
Db 181 ALDLYPYDAGTDSGFTSSPNFATIPQDVTETITSSPSHPANSFYYPRLKALPPIARVT 240
Qy 241 LVRLRQSPRAFIPAPVLPSPRDNEIVDSASVPETPLDCEVSLWSSWGLCGHCGRLGTSK 300
Db 241 LLRLRQSPRAFIPAPVLPSPRDNEIVDSASVPETPLDCEVSLWSSWGLCGHCGRLGTSK 300
Qy 301 RTRVYRVQPNANGSPCPPELEEEACVPDNCV 331
Db 301 RTRVYRVQPNANGSPCPPELEEEACVPDNCV 331

RESULT 3

Q9W75 PRELIMINARY; PRT; 330 AA.
ID Q9W75;
AC Q9W75;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MINDIN PRECURSOR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE DAWLEY;
RA Feinstein Y.; Borrell V.; Garcia C.; Burstyn-Cohen T.; Tzarfaty V.;
RA Frunkin A.; Nose A.; Okamoto H.; Higashijima S.; Soriano A.; Klar A.;
RT "F-spondin and mindin: two structurally and functionally related genes
expressed in the hippocampus that promote outgrowth of embryonic
hippocampal neurons.";
RL Development 0:0-0(1999).
DR EMBL; AF155196; AAD38195.1;
DR InterPro; IPR000884; TSPL.
DR Pfam; PF00090; tsp_1; 1.
DR PROSITE; PS50092; TSPL; 1.
DR SMART; SM00209; TSPL; 1.
KW Signal.
FT SIGNAL. 1 22 POTENTIAL.
ET CHAIN 23 330 MINDIN.
SQ SEQUENCE 330 AA; 36014 MW; ECBCF07A0345A83A CRC64;

Query Match 86.6%; Score 1520.5; DB 11; Length 330;
Best Local Similarity 86.4%; Pred. No. 1.3e-120;
Matches 287; Conservative 17; Mismatches 25; Indels 3; Gaps 2;

Qy 1 MENPSAALGKALLATLGAAGPLGSESTCSARALAKYSITITGKWSQTAPFKQY 59
Db 1 MENVS--FSLDRTLWVFLMLAGTAGQPLGSESVCTARPLARYSITITGKWSQTAPFKQY 58
Qy 60 YPLFRPPAOWSSLLGAHSSDYSMRKNQYVSNGLRDFAEERGEAWALMKEIEAAGEALQSV 119
Db 59 YPLFRPPAOWSSLLGAHSSDYSMRKNQYVSNGLRDFAEERGEAWALMKEIEAAGEALQSV 118
Qy 120 HAVFSAPAVPGTGTSAELEVQRHSLVSFVVRIVSPDWFVGVDSLDLDCGDRWREQ 179
Db 119 HAVFSAPAVPGTGTSAELEVQRHSLVSFVVRIVSPDWFVGVDSLDLDCGDRWREQ 178
Qy 180 ALDLYPYDAGTDSGFTSSPNFATIPQDVTETITSSPSHPANSFYYPRLKALPPIARV 239
Db 179 VLDLYPHDAGTDSGFTSSPNFATIPQDVTETITASSPSHPANSFYYPRLKSLPIAKV 238
Qy 240 TLVRLRQSPRAFIPAPVLPSPRDNEIVDSASVPETPLDCEVSLWSSWGLCGHCGRLGTSK 299
Db 239 TFVRLRQSPRAFIPAPVLPSPRDNEIVDSASVPETPLDCEVSLWSSWGLCGHCGRLGTSK 298
Qy 300 SRTRYRVQPNANGSPCPPELEEEACVPDNCV 331
Db 299 SRTRYRVQPNANGSPCPPELEEEACVPDNCV 330

RESULT 4

Q42112 PRELIMINARY; PRT; 331 AA.
ID Q42112;
AC Q42112;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MINDIN2.
OS Brachydanio rerio (Zebrafish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Higashijima S.; Nose A.; Eguchi G.; Hotta Y.; Okamoto H.;
RL Dev. Biol. 0:0-0(1997).
DR EMBL; AB006085; BAA22809.1;

DR InterPro: IPR000884; TSPL.
DR Pfam: PF00090; tsp_1; 1.
DR PROSITE: PS0092; TSPL; 1.
DR SMART: SM00209; TSPL; 1.
SQ SEQUENCE 331 AA; 36887 MW; 1D95D82B6549D273 CRC64;

Query Match 63.1%; Score 1108; DB 13; Length 331;
Best Local Similarity 61.6%; Pred. No. 9.3e-86;
Matches 197; Conservative 51; Mismatches 64; Indels 8; Gaps 3;

QY 17 LLATL- GAAGOLPGGESICSRALAKYSITFTGKWSQTAFPKOYPLFRPPAQSLLGA 75
DB 15 MTALLGVAMPVDVDMCTASTAKYLTFTGWTQTAFAKHYLYRPPAQSLLGV 74

QY 76 AHSDYSWMRKNOYVNSGLRDFAEERGEAWALMKEIEFAAGALQSVAHVFSAPVPGTGQ 135
DB 75 THSSDYHLQORNEYASNGVREFSERAEAWTLKEVEAAGERIQSVYGLFSAPAVAGTGH 134

136 TSABLEVQRHSLVSFVVRIVPSPDFVGVDSLDLDCGDRWREQAALDLYPYDAGTDSGF 195
DB 135 ATTEFEVFAHSLLSFTVRIVPSPDNFVGVDSNLNLCGDKENISLEULPYDAGTDSGF 194

QY 196 TSSPNFATIPQDVTEITSSSPSHANSFYPRKALPPIARVTLVRLRQSPRAFIPPA 255
DB 195 TFSNPFETIPQDKVTOITSSFSHPANSFYPRKHLPIAKVSLTKIKNNQ--IFSL 251

QY 256 PVLPSRNEIVDSASVP-----ETPLDCEVSLWSWGLCGHGRGLGKTSRTRVVRVOPAN 311
DB 252 PIQQTQSNQIPSGNEIDGDLINTPLDCEVSWSPWGLCKGCGEKGVKHKRTRYIHMHPAN 311

QY 312 NGSPCELEEEAECPDNCV 331
DB 312 NGAPCPSLEERKLCIPDNCV 331

RESULT 5
042111 ID O42111 PRELIMINARY; PRT; 334 AA.
AC O42111;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MINDIN.
GN MINDIN2 OR MINDIN1.
OS Brachydanio rerio (zebrafish) (Zebra daniel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi.
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
NCBI_TaxID=7955;
[1]
SEQUENCE FROM N.A.
RA Higashijima S., Nose A., Eguchi G., Hotta Y., Okamoto H.;
RL Dev. Biol. 0:0-0(1997).
DR EMBL: AB006084; BAA22808.1; -.
DR ZFIN: ZDB-GENE-990415-160; mindin1.
DR ZFIN: ZDB-GENE-990415-161; mindin2.
DR InterPro: IPR000884; TSPL.
DR Pfam: PF00090; tsp_1; 1.
DR PROSITE: PS0092; TSPL; 1.
DR SMART: SM00209; TSPL; 1.
SQ SEQUENCE 334 AA; 37233 MW; 7451BF2F95AEDF05 CRC64;

Query Match 49.4%; Score 868; DB 13; Length 334;
Best Local Similarity 49.7%; Pred. No. 1.8e-65;
Matches 166; Conservative 51; Mismatches 91; Indels 26; Gaps 6;

QY 10 LCKALCALATLTLGA--GOPLGSGESICSRALAKYSITFTGKWSQTAFPKOYPLFRPP 66
DB 12 LQOLLVWLRFTLSAALVNSTNGTE--CSARGPASYIVVFTGHSWSPOTFFPKOYPLFRPP 69

QY 67 AQWSSILGAHSSDYSWMRKNOYVNSGLRDFAEERGEAWALMKEIEFAAGALQSVAHVFS 126

DB 70 AQWSKLMVVTTHNEQYRLWQEGAPASDGMKSFAGQGLTVDLVKDAKARKR-RSVGSMYRT 128
QY 127 PAVPSGTQGTSAELVQRHSLVSFVVRIVPSPDFVGVDSLDLDCGDRWREQAALDLYP 186
DB 129 AGIPSGIGHSSSTEVLLTPRSPVLVSLVLPSPDFVGVDSNLNLCGDKENISLEULPYDAGTDSGF 188

QY 187 YDAGTDSGFTSSPNFATIPQDVTEITSSSPSHANSFYPRKALPPIARVTLVLR--- 243
DB 189 FDAGTDSGFTSSPNFATIPQDVTEITSSSPSHANSFYPRKALPPIARVTLVLR--- 248

QY 244 --LRQSPRAFIPPAVLPVPSRNEIVDSASVP-----ETPLDCEVSLWSWGLCGHGRGLG 297
DB 249 LPVROQNRL-----SNHILPDASKPHRFSETPDCEVSWSPWGLCGHGRGLG 297

QY 298 TKSRTYRVVQVQANNSPCPELEEEAECPDNCV 331
DB 298 LRHRTYRILLKPNANSGPCPELEEEAECPDNCV 331

RESULT 6
09V746 ID Q9V746 PRELIMINARY; PRT; 601 AA.
AC Q9V746;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MSPO PROTEIN
GN MSPO OR CG10145.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H.C., Blaise J.R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Bianknoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Durbin K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fasker C., Fasker A.E., Garb N.S., Gelbart W.M., Glaser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Haris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.M., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo R., Pittman G.S., Pan S., Pollard J., Pur V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
[1]

Q9V746 ID Q9V746 PRELIMINARY; PRT; 601 AA.
AC Q9V746;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MSPO PROTEIN
GN MSPO OR CG10145.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H.C., Blaise J.R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Bianknoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Durbin K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fasker C., Fasker A.E., Garb N.S., Gelbart W.M., Glaser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Haris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.M., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo R., Pittman G.S., Pan S., Pollard J., Pur V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
[1]

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195 (2000).
DR EMBL; AE003813; AAF58219.1; -.
DR FlyBase; FBgn0020269; msp0.
DR InterPro; IPR002106; AA.TRNA_ligase_II.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00090; Tsp_1; 1.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS00339; AA.TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS5092; TSPI; 1.
SQ SEQUENCE 601 AA; 65434 MW; 529FEDCC348AC8FF CRO

Query Match 30.1%; Score 528.5; DB 5; Length 601;
Best Local Similarity 27.2%; Pred. No. 1.8e-36;

12:

Qy		2	ENFSPAAALGKALCALLATLGAAGOPGLGESICSARALAKYSITFTGWKSQTAFPKQP	61
			: : : : :	
		86	DDSPVSNLPTAS--LAT-PPATQP-APOTCTLDRLAVYKVVLHTYTRELFPKHYP	140
			: : : : :	
Qy		62	LFRPPAQWSSILGAHSSDYSMRKNQOYVNSGLRDFAEERGEA-----WALM	107
			: : : : :	
Db		141	DWRPTAQWTLTGRTHNANTALHYIQPATAAVKQFAESGRTDLLDSNAGEQQOVOMQJQ	200
			: : : : :	
Qy		108	KEIEAAGE-----ALQSVAHFVS	125
			: : :	
Db		201	SOMQAOKSPSGGISGTTSFNATAASTATPTGGSGSGSGGGTGTTTAERSVPDEFS	260
			: : :	
Qy		126	APAVPSGTGOTSAELEVQRHSLSVFWIRIVSPDWFGVDSLDCDGRWREQAALDIY	185
			: : : : :	
Db		261	MPAIPMGAGSEAKVFVDSNHLSLVLMTRIVSPDMFIGCVSFLCWSGSWIDTVTVELD	320
			: : : : :	
Qy		186	PYDAGTDSGTFESSPNFATIPQDTVEITSSPSHPANSFYYPRLKALPIIARVTLVRL-	244
			: : : : :	
Db		321	PLDAGTDNGFTFTAPNWPEPGQVIYRIITSRYPGHPAGSFYYPKSKRUPTATTFQIKL	380
			: : : : :	
Qy		245	-----244	
			: : : : :	
Db		381	EYELSEVNIAEDDRKYETVQOTHLDAEHNVENNELSAISIEREROTEQOOLQONDE	440
			: : : : :	
Qy		245	RQSPRA-----FIPPAP-----VLPSRDNEIV-----DSAS	270
			: : :	
Db		441	RQIRSOALLAKMNPIYGSNNSLOPAGQVVSVVPKNDXHA ILQSTASSYRRAADASANA	500
			: : : : :	
Qy		271	VPETP-----LDCEVSLWSWGICGGHCGRLGTKGSRTR	303
			: : : : :	
		501	SRPTPSAIGGGKAGGAVGGGAATRRRSSAQRNRDCRVSHSWETAACKSCG-VGEMHIYR	559
			: : : : :	
Qy		304	YVRVOPANNGPSCPLEEEBAECVPD-NC	330
			: : : : :	
Db		560	KVIKKRGGRGROCPALOOKSKWCCTERNC	587
			: : : : :	

DECEMBER 7

RESULT /
002029
ID 002029
PRELIMINARY;
PRT: 598 AA.

AC	O02029;
DC	MSPD OR C010145.
DT	Drosophila melanogaster (Fruit fly).
DD	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda
DT	Pterygota; Neoptera; Endopterygota; Diptera; Brachy
DE	Ephyridioidea; Drosophilidae; Drosophila.
OC	NCBI_TaxID=7227;
OX	

[1]
SEQUENCE FROM N.A.
STRAIN=CANTON S;
Umemiya T., Takeichi M., Nose A.;

RL	Dev. Biol.	0:0-0(0).	
DR	EMBL	AF000178; AAB38341.1; -.	
DR	FlyBase	FBgn020269; mspro.	
DR	InterPro	IPR002106; AA:trna_ligase_II.	
DR	InterPro	IPR000884; TSP1.	
DR	Pfam	PF00090; tsp1; 1.	
DR	SMART	SM00209; TSP1; 1.	
DR	PROSITE	PS00339; AA:TRNA_LIGASE_II_2; 1.	
DR	PROSITE	PS50092; TSP1; 1.	
SQ	SEQUENCE	598 AA; 65251 MW; 02973A52	

Query Match 29.5%; Score 518; DB 5; Length 598;
Best Local Similarity 26.5%; Pred. No. 1.4e-35;

11;

QY	2	ENPSPAALGKALCALLATILGAAGOPLGSESTCSARALAKYSITFTGKWSOTAFPKQYP	61
Db	86	DDPSPVNNPT--ASLATPATQATPQTG--CTLDRLAVTKVVLHTYWTRELFPKHP	140
QY	62	LFRPPAOWSSLLGAHSDYSWRKKQYVSNGLRQFAERGEA-----WALM	107
Db	141	DWRPTAQWTKILGRTHNANTALIHIGOPATAAVKQFAESGRTDLLDSNAGQQOQVOMQLQ	200
QY	108	KTEIAGE-----ALOSVHIAVFSAPA	128
Db	201	SOMQAGKSPSGISSGTTSFNATAASTATPTGGSGSGSGVTGTTTAERSVPDEFSPMA	260
QY	129	VPSGTFGOTSAGEVORRHSIVSVFVRLVPSDQWFGVDSLDICDGRWRQOALDLYPYD	188
Db	261	IPMGAGREAOVFVDSNHSLSVLSMTIRVSPDWFIGVDSFELCVGGSWIDTVTVTDLPLD	320
QY	189	ACTDGSFTFSSPNFATIPQDTVTETITSSPSHPANSFYPRKALPIARVTLVRL----	244
Db	321	ACTDNGFTFTAPNNPTAQOGVIYRIUTSRYPGHPAGSFYYPKSKRLPIATFQFIKLKEYE	380
QY	245	-----ROS	247
Db	381	LSEVENTAEDDRKYETVOTQTHLDAEHNIVEMNNELASIERERQTHQOOLQONDDERQ	440
QY	248	PA-----FIPPAP-----VLPSRDNEIV-----DSASVPE	273
Db	441	IRSOQLAKMNPYGSNNLSQAAAPQGVSVVPKNDKHAILQSIASSYRRAADASDANASPK	500
QY	274	Tp-----LDCVSLWSNWGLCGCHCGCRGLGTTKSRTRYVR	306
Db	501	TPSAIGGGKAGGAVGGGAATRRRSNORRDCRVSHNSEWTACKSKCG--VGMHRYRKVI	559
QY	307	VOPANNSPCELEEEAECPD--NC	330
Db	560	KHGRGRGRCQPALQOQKWCGRTERNC	584

RESULT

RESULTS 6
O42113
ID O42113 PRELIMINARY; PRT: 808 AA.

AC	O42113;	
AD	01-JAN-1998 (TRENBLrel. 05, Created)	
DT	01-JAN-1998 (TRENBLrel. 05, Last sequence update)	
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)	
DE	F-SPENDINI.	
OS	Brachydanio rerio (Zebrafish) (Zebra danio).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;	
OC	Cypriniformes; Cyprinidae; Rasbora; Rasbora; Danio.	
OX	NCBI_TaxID=7955;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RL	Higashijima S., Nose A., Eguchi G., Hotta Y., Okamoto H.;	
RA	Dev. Biol. 0:0-0(1997).	
DR	EMBL; AB006086; BAA22810.1; -.	
DR	InterPro; IPR000884; TSPI.	
DR	InterPro; IPR002861; Reeler.	

DR Pfam: PF00090; tsp.1; 6.
 DR Pfam: PF02014; Reeler; 1.
 DR PROSITE: PS00092; TSP1; 6.
 DR SMART: SM00209; TSP1; 6.
 SQ SEQUENCE 808 AA; 90645 MW; 0A24154AA4A89EC7 CRC64;

Query Match 26.8%; Score 470.5; DB 13; Length 808;
 Best Local Similarity 33.2%; Pred. No. 2.1e-31;
 Matches 114; Conservative 52; Mismatches 132; Indels 45; Gaps 11;

QY 9 ALGKALCALLATLGAAGQPLGGESIT-----CSARALAKYSITFTGKWSQTAFPKQYPL 62
 Db 176 SLTRKMC-----EKESLYGETTKPLDCCAGCTAKYRTVYGNWSEKHLPKDYP- 225
 QY 63 FRPPAQWSSILGAHSDYSNMRKNQVSNGLRDFARGEAWALMKEIEAAGEALQSVHA 122
 Db 226 -RANHSALIGASHSKNLYWEYGYASEGVQVAGELGSPVKMEEIRQKQDEVLTIVK 284
 123 V-----FSAPVPSGTGTSAELEVQRHSLVSFVVRIVPSPDMFVGVDSLDLDCG 173
 285 MKAQWPAWQPLNVAAP-----SAEFSVDRTRHLMSEFLTMLGSPDMNVGLSSEDLCTR 338
 QY 174 D-RWRQAAALDLYPDAGTSGTSSPNFATIPQDVTVEITSSSPSHANSFYPRLKA 232
 Db 339 ECGWQKVQVQDLIPDAGTSGVSESPNKSAPQEKIRPLTSL--DHPQSPFYDPGGA 396
 QY 233 LPPIARVTLVRL-RQSPRAFIPAPVLPSPKNEIVDSA----SVPETPLDCEVSLWSWG 287
 Db 397 IFPVALVVERTARKEQCNIPVDNV---DQIVADIAQEEEDTPTETCIYSNWSWPS 452
 QY 288 LCGHGCGRLGTSRTRYRVQPNANNSPCPELEEEAECPDNC 330
 Db 453 ACSSTCEKGRMRORMLKAQ-LDLSVPCPHTQDPEPCIGGCC 494

RESULT 9
 Q9W770 PRELIMINARY; PRT; 802 AA.
 AC Q9W770;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DE F-SPONDIN PRECURSOR.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 RN Gallus.
 NCBI_TaxID=9031;
 [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99211385; PubMed=10197528;
 RX Debby-Brafman A., Burstin-Cohen T., Klar A., Kalcheim C.;
 RT "F-spondin, expressed in somite regions avoided by neural crest cells,
 RT mediates inhibition of distinct somite domains to neural crest
 RT migration.";
 RL Neuron 22:475-488(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99325518; PubMed=10399931;
 RA Burstin-Cohen T., Tzarfaty V., Frumkin A., Feinstein Y., Stoeckli E.,
 RA Klar A.;
 RT "F-spondin is required for accurate pathfinding of commissural axons
 RT at the floor plate.";
 RL Neuron 23:233-246(1999).
 DR EMBL: AF149302; RAD41495.1; .
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR002861; Reeler.
 DR Pfam: PF00090; tsp.1; 6.
 DR Pfam: PF02014; Reeler; 1.
 DR PROSITE: PS00092; TSP1; 5.
 DR SMART: SM00209; TSP1; 6.
 KW Signal.

FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 802 F-SPONDIN.
 SQ SEQUENCE 802 AA; 90509 MW; 0644D2BDD0A0FE12 CRC64;

Query Match 26.5%; Score 464.5; DB 13; Length 802;
 Best Local Similarity 33.2%; Pred. No. 6.8e-31;
 Matches 111; Conservative 52; Mismatches 144; Indels 27; Gaps 10;

QY 9 ALGKALCALLATLGAAGQPLGGESIGSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
 Db 171 SLTRICEQDSASGVTDKP---TLDCACGCTAKYRLTFYGNWSEKTHPKDFP--RRTNH 225
 QY 69 WSSLLGAHSDYSNMRKNQVSNGLRDFARGEAWALMKEI-EAAGEALQSVHAVESAP 127
 Db 226 WSAITGSHSKNLYLWEYGYASEGVQVAGELGSPVKMEEIRQDSDEVLTIVKAKAQPW 285
 QY 128 AVP--SGTGQTSAELEVQRHSLVSFVVRIVPSPDMFVGVDSLDLDCGD-RWREQAALDL 184
 Db 286 AWQPLNVAAPSAEFSVDRTRHLMSEFLTMLGSPDMNVGLSAEDLCTKDCGWQKVQVQDL 345
 QY 185 YPDAGTSGTSSPNFATIPQDVTVEITSSSPSHANSFYPRLKALPPIARVTLVRL 244
 Db 346 IPWDAGTSGVSESPNKSAPQEKIRPLTSL--DHPQSPFYDPGGSIKIVARVLERI 403
 QY 245 RQSPRA--FIPP-----APVLPSPKNEIVDSASVPETPLDCEVSLWSWGLCGHCGRL 296
 Db 404 ARKGEQCNFVNDIDIVADLAPEKEE-----DDTPTETCIYSNWSWPSACSSTCEK 456
 QY 297 GTRSTRVVRVQPNANNSPCPELEEEAECPDNC 330
 Db 457 GKRMRMLKAQ-LDLSVPCPHTQDPEPCMGPGC 489

RESULT 10
 Q9HCB6 PRELIMINARY; PRT; 807 AA.
 AC Q9HCB6;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VSGP/F-SPONDIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RA Miyamoto K., Morishita Y., Yamazaki M., Minamino N., Kangawa K.,
 RA Matsuo H., Mizutani T., Yamada K., Minegishi T.;
 RT "Isolation and characterization of vascular smooth muscle cell growth
 RT promoting factor from bovine ovarian follicular fluid, and its CDNA
 RT cloning from bovine and human ovary.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB051390; BAB18461.1; .
 DR InterPro: IPR002861; Reeler.
 DR InterPro: IPR000884; TSP1.
 DR Pfam: PF00090; tsp.1; 6.
 DR Pfam: PF02014; Reeler; 1.
 DR SMART: SM00209; TSP1; 6.
 DR PROSITE: PS00092; TSP1; 5.
 SQ SEQUENCE 807 AA; 90987 MW; 06FC52484206BB55 CRC64;

Query Match 26.2%; Score 460.5; DB 4; Length 807;
 Best Local Similarity 34.5%; Pred. No. 1.5e-30;
 Matches 115; Conservative 50; Mismatches 143; Indels 25; Gaps 11;

QY 9 ALGKALCALLATLGAAGQPLGGESIGSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
 Db 176 SLTKLCEQDSTFDGVTDKPI---LDCCACGCTAKYRLTFYGNWSEKTHPKDYP--RRANH 230
 QY 59 WSSLLGAHSDYSNMRKNQVSNGLRDFARGEAWALMKEI-EAAGEALQSVHAVFSAP 127

```

Db 231 WSAIIGGSHSNVYVWYGGVASEGVKQVAGELGSPVKKMEIEIROQSDSEVLTVIKAKAQP 290
QY 128 AVP--SGTGQTSAELEVORRHSLVSFVVRIVPSPDFWGVDSLDLDCGD--RWREQAALDL 184
Db 291 AWQPLNVAAPSAEFSVDRTHLSFMTMGPSPDWNVGLSAEDLCTKEGWNQVQVQDL 350
QY 185 YPDAGTSGTFFSPNPATIPQDTVTETSSSPSHSPANSFYPRKALPPIARVTLVRL 244
Db 351 IPWDAGTSGVTYESPNKPTIPQEKIRPLTSL--DHPQSPFYDPGGISITQVARVIERI 408
QY 245 -RQSPRAPIPPAPVLPDRNEIVDSASVPE-----TPDCEVSLWSWGLCGHCHGRIG 297
Db 409 ARKGQCQIVPDNV-----DDIVADLA--PEEKDEDDTPTETCIYSNNWSPWACSSTCDKG 462
QY 298 TKSRTRYVRVQPANNGSPCPELEAEACVPCNC 330
Db 463 KMRQRMKLAQ-LDLSVPCPDQDFQPCMGPGC 494

RESULT 11
ID 09GLX9 PRELIMINARY; PRT; 807 AA.
AC 09GLX9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-JUN-2001 (Tremblrel. 16, Last sequence update)
DE VSGP/F-SPONDIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyamoto K., Morishita Y., Yamazaki M., Minamino N., Kangawa K.,
RA Matsuo H., Mizutani T., Yamada K., Minegishi T.;
RT "Isolation and characterization of vascular smooth muscle cell growth
RT promoting factor from bovine ovarian follicular fluid, and its cDNA
RT cloning from bovine and human ovary."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB051389; BAB18460.1; -.
DR InterPro; IPR002861; Reeler.
DR InterPro; IPR000884; TSPL.
DR Pfam; PF00090; Reeler; 1.
DR Pfam; PF00090; tsp_1; 6.
DR SMART; SM00209; TSPL; 6.
DR PROSITE; PS50092; TSPL; 5.
DR SEQUENCE 807 AA; 90976 MW; 4C484B331FB1034C CRC64;

Query Match 26.2%; Score 460.5; DB 6; Length 807;
Best Local Similarity 34.5%; Pred. No. 1.5e-30;
Matches 115; Conservative 50; Mismatches 143; Indels 25; Gaps 11;

QY 9 ALGRKALLATLGAQGPUGGSGISARALAKYSITFTGKWSQTAFFKQYPLFRPPAQ 68
Db 176 SLTKKLCEQDSTFGVDTKPI--LDCCAGCTAKYRLTFYGNWSEKTHPKDYP--RRAN 230
QY 69 WSSLGAAHSDYSWMRKNQVSNGLRDFARGEAWALMKEI-EAAGEALQSVHAFVSAP 127
Db 231 WSAIIGGSHSNVYVWYGGVASEGVKQVAGELGSPVKKMEIEIROQSDSEVLTVIKAKAQP 290
QY 128 AVP--SGTGQTSAELEVORRHSLVSFVVRIVPSPDFWGVDSLDLDCGD--RWREQAALDL 184
Db 291 AWQPLNVAAPSAEFSVDRTHLSFMTMGPSPDWNVGLSAEDLCTKEGWNQVQVQDL 350
QY 185 YPDAGTSGTFFSPNPATIPQDTVTETSSSPSHSPANSFYPRKALPPIARVTLVRL 244
Db 351 IPWDAGTSGVTYESPNKPTIPQEKIRPLTSL--DHPQSPFYDPGGISITQVARVIERI 408
QY 245 -RQSPRAPIPPAPVLPDRNEIVDSASVPE-----TPDCEVSLWSWGLCGHCHGRIG 297

```

```

Db 409 ARKGQCQIVPDNV-----DDIVADLA--PEEKDEDDTPTETCIYSNNWSPWACSSTCDKG 462
QY 298 TKSRTRYVRVQPANNGSPCPELEAEACVPCNC 330
Db 463 KMRQRMKLAQ-LDLSVPCPDQDFQPCMGPGC 494

RESULT 12
ID 094862 PRELIMINARY; PRT; 624 AA.
AC 094862;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE KIAA0762 PROTEIN (FRAGMENT).
GN KIAA0762.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 5:277-286(1998).
DR EMBL; AB018305; BAA34482.1; -.
DR InterPro; IPR000884; TSPL.
DR Pfam; PF00090; tsp_1; 6.
DR PROSITE; PS50092; TSPL; 5.
DR SMART; SM00209; TSPL; 6.
DR NON_TER 1
FT NON_SEQUENCE 624 AA; 70557 MW; 40F2238D29024D03 CRC64;

Query Match 26.0%; Score 457; DB 4; Length 624;
Best Local Similarity 35.5%; Pred. No. 2.1e-30;
Matches 109; Conservative 47; Mismatches 129; Indels 22; Gaps 10;

QY 35 CSARALAKYSITFTGKWSQTAFFKQYPLFRPPAQSSLLGAHSDYSWMRKNQVSNGL 94
Db 16 CCACGTAKYRLTFYGNWSEKTHPKDYP--RRANHSATIGGSHSNVYVWYGGVASEGV 73
QY 95 RDFARGEAWALMKEI-EAAGEALQSVHAFVSAPVP--SGTGQTSAELEVORRHSLVSF 151
Db 74 KQVAELGSPVKMEIEIROQSDSEVLTVIKAKAQPAPQPLNVAAPSAEFSVDRTHLSMF 133
QY 152 VVRIVPSPDFWGVDSLDLDCGD--RWREQAALDLYPDAGTSGTFFSPNPATIPQDTV 210
Db 134 LTMWGPSDWNVGLSADLCTKEGWNQVQVQDLIPWDAGTSGVTYESPNKPTIPQEKI 193
QY 211 TETSSSPSHSPANSFYPRKALPPIARVTLVRL-RQSPRAPIPPAPVLPDRNEIVDSA 269
Db 194 RLPLTSL--DHPQSPFYDPGGISITQVARVIERIARKGECQCNVPDNV---DDIVADLA 247
QY 270 SYPE-----TPDCEVSLWSWGLCGHCHGRIGLTKSRTRYVRVQPANNGSPCPELEBEA 323
Db 248 --PEEKDEDDTPTETCIYSNNWSPWACSSTCDKGKMRQRMKLAQ-LDLSVPCPDQDFQ 304
QY 324 ECVPCNC 330
Db 305 PCMGPGC 311

RESULT 13
ID 076822 PRELIMINARY; PRT; 898 AA.
AC 076822;
DT 01-NOV-1998 (Tremblrel. 08, Created)

```


DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS50092; TSPI; 4.
KW Serine protease inhibitor.
SQ SEQUENCE 763 AA; 84945 MW; 3292

Query Match	25.4%	Score 445.5	DB 5	Length 763
Best Local Similarity	33.3%	Pred. No. 2.6e-29		
Matches 111; Conservative	45	Mismatches 138	Indels 39	Gaps 9

Oy 2 ENPSPAAALGKALCALLATLGAAGPLGGESICSRALAKYISITFTGKWSTAFPKQYP 61
 | ||| | | | | | | | | |
Db 173 ERKPDAAAQAQKECCA-----CDE----AKYSEVFEGTWSNETHPKDYP 211

Qy 62 LFRPAQWSSLLGAHSSDYSMWKRNQYVNSNGLRDFAERGEAWALMKEIEAGEALQSVH 121

QY 122 AVFSAVAV--PSGTGQTSAELEVQRHSLVSFVVRIVPSPDFWVGVDSDLDCDGD-RWRE 178

270 -LIKAAGLWYPNVNQNTSSKFRVDRKHKPVSLVSMFGSPDWVVGISGLDLCTEDCSNKE 328

D**b** 329 SMDFLPWDAGTDSGISYMSNPSETQPPERMYRITTMYPEDPRAPFYNPKSREMTPLAK 388

Qy 239 VTLVR----LRQSPRAFIAPPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCG 294
:
:
:
Dh 389 LVLPBKEITVSNCDNET-----LGMQIETVEKAPPPGPAEPGVCSFAVGLPQQYVQLA 443
:
:
:
:

QY 295 RLGTKSRTR-YVRVQPANNGSPCELEEEAECV 326
 : :

Db 443 K-GIRMSRQVLYPAAADQNKARQLVAKENCV 474

RESULT 15

ID	PRELIMINARY	PRT	803 AA.
042114			
AC	042114		
DT	01-JAN-1998 (TReMBLrel. 05, Created)		
DT	01-JAN-1998 (TReMBLrel. 05, Last sequence update)		
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)		
DE	F-SPONDIN2		
OC	Brachydanio rerio (zebrafish) (Zebra danio).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;		
OC	Cypriniformes; Cyprinidae; Rasbora; Danio.		
	NCBI_TaxID=7955;		

[1]

DR	SEQUENCE FROM N.A.
DR	Higashijima S., Nose A., Eguchi G., Hotta Y., Okamoto H.;
RL	Dev. Biol. 0:0-0(1997).
DR	EMBL; AB006087; BAA32811.1;
DR	InterPro; IPR000762; PTN_MK.
DR	InterPro; IPR000884; TSP1.
DR	InterPro; IPR002861; Reeler.
DR	Pfam; PF00090; tsp_1; 6.
DR	Pfam; PF02014; Reeler; 1.
DR	PROSITE; PS50092; TSP1; 6.
DR	SMART; SM00193; PTN; 1.
DR	SMART; SM00209; TSP1; 6.
DR	SEQUENCE 803 AA: 90331 MW: 619C8F5A840DC158 CRC64:

Query Match	25.2%	Score 443;	DB 13;	Length 803;
Best Local Similarity	30.6%	Pred. NO. 4.5e-29;		
Matches 113; Conservative	47;	Mismatches 131;	Indels 78;	Gaps 10;

35 CSARALAKYSITFTGKWSOTAFPKQYPLFRPPAQWSSLLGAHSSDYSMWRKKNQYVSNGL 94

Dbb 194 CCACGTAKYRVTFYGNWSEKLPKDYP--RRANHWSALIGASHTRSYVLWEYGGFASEGV 251

Qy	95	RDFERGEAWLMKEIEAAGALQSV---HAVFSAPAVPSGQTGQTAELVQRHSLVSF	151
Db	252	KQVSEYSPVKMEIEIRQKGGDVMYTIKTKAOWPAWQPLNIRAAPSAEFSVDVRHLISF	311
Qy	152	VVRIVPSPDWFGVDSLUDLDCGD-RWRQEAALDLYPYDACTDSGETFSSPNFATIPQDTV	210
Db	312	LTMLGSPDNNVGLSGEDLCTRCGWQVKLVKDLVPWDAGTDSGVSEPNKPSIPQERI	371
Qy	211	TEITSSPSPIHANSFYYPULKALPIIARVTLVRLRQ-----	246
Db	372	RLPLTSL--DHQPSPFYDPSPGGPMTPLGRVVVVERIARKGEOCNTPVPTIDDIVADIAQEEK	429
Qy	247	-----SPRAFI-----PPAVLPSPRNE-IVDSA	269
Db	430	EEGDNTPETCIYSNWSPWSACSSSTCDKGRMRQRMKALQDLNPLCLHTQDFEPCMGPG	489
Qy	270	SVLPETPLDCVSLWSWGLCGHGCRIGTKSRTRYRVYVQFPAWNGSPCPLEBEAE	324
Db	490	CSBEEASTCMMEIWSWSPSCASCG-MGMRSRERYVKQFP-EDGSSCTLPTTEETECVVN	547
Qy	325	--CVPDNCV	331
Db	548	DDCSPSSCV	556

Search completed: December 10, 2001, 20:02:39
Job time: 2891 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2001, 13:37:26 ; Search time 1640.62 Seconds
(without alignments)
11111.277 Million cell updates/sec

Title: US-09-170-042A-1

Perfect score: 1105

Sequence: 1 cgtgtctctgcgggtgat.....tgcaggcgccgaggcgacaca 1105

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vl.*

30: em_htgo_hum.*

31: em_htgo_inv.*

32: em_htgo_rod.*

33: em_htg_hum.*

34: em_htg_inv.*

35: em_htg_rod.*

36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1105	100.0	1105	6	AR035961	AR035961 Sequence
2	1098.6	99.4	1634	9	BC002707	BC002707 Homo sapi
3	1084.8	98.2	1807	9	AB027466	AB027466 Homo sapi
4	665.6	60.2	2325	10	AF155196	AF155196 Rattus no
5	362.6	32.8	2116	5	AB006085	AB006085 Danio rer
6	343.8	31.1	400	6	AR037874	AR037874 Sequence
7	339	30.7	400	6	AR125230	AR125230 Sequence
8	336.2	30.4	197176	2	AC022763	AC022763 Homo sapi
9	300	27.1	506	6	AR035966	AR035966 Sequence
10	279	25.2	4292	9	AK024499	AK024499 Homo sapi
11	277.4	25.1	2820	9	AK026054	AK026054 Homo sapi
12	258	23.3	3381	5	AB006084	AB006084 Danio rer
13	202.6	18.3	316	6	AR035967	AR035967 Sequence
14	202.6	18.3	316	6	AR035968	AR035968 Sequence
15	118.2	10.7	2585	3	DMAF000178	AF000178 Drosophill
16	110.6	10.0	3489	3	BFL6096	AJ006096 Branchios
17	105	9.5	2466	3	AF135119	AF135119 Drosophill
18	81.2	7.3	3146	5	AB006086	AB006086 Danio rer
19	78	7.1	4061	4	AB051389	AB051389 Bos tauru
20	78	7.1	4061	6	E32471	E32471 Novel vascu
21	75	6.8	51289	3	AC005717	AC005717 Drosophill
22	75	6.8	73721	2	AC020007	AC020007 Drosophill
23	75	6.8	259817	3	AE003804	AE003804 Drosophill
24	71.6	6.5	3079	6	AX093366	AX093366 Sequence
25	71.6	6.5	3079	6	E32470	E32470 Novel vascu
26	71.6	6.5	3079	9	AB051390	AB051390 Homo sapi
27	71.6	6.5	3999	6	AX093359	AX093359 Sequence
28	71.6	6.5	3999	9	AB018305	AB018305 Homo sapi
29	70.6	6.4	76217	2	AC020385	AC020385 Drosophill
30	70.6	6.4	160903	3	AC007580	AC007580 Drosophill
31	70.6	6.4	258778	3	AE003813	AE003813 Drosophill
32	70	6.3	4029	6	AR007441	AR007441 Sequence
33	70	6.3	4029	6	E32472	E32472 Novel vascu
34	70	6.3	4029	10	RATFSAA	M88469 Rattus norv
35	67	6.1	20293	1	SCD20	AL392148 Streptomy
36	65.4	5.9	10805	2	AC017707	AC017707 Drosophill
37	65.4	5.9	184621	3	AC007440	AC007440 Drosophill
38	65.4	5.9	266133	3	AE003822	AE003822 Drosophill
39	61.6	5.6	3226	5	AF149302	AF149302 Gallus ga
40	61.6	5.6	3226	6	AR007442	AR007442 Sequence
41	61.2	5.5	59843	8	AP000816	AP000816 Oryza sat
42	61.2	5.5	161266	8	AP001278	AP001278 Oryza sat
43	60.8	5.5	12095	1	AE005853	AE005853 Caulobact
44	60.8	5.5	167761	2	AP003221	AP003221 Oryza sat
45	60	5.4	3576	5	AB006087	AB006087 Danio rer

ALIGNMENTS

RESULT 1

AR035961

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

BASE COUNT

ORIGIN

PAT 29-SEP-1999

AR035961 1105 bp DNA
Sequence 1 from patent US 5871969.

AR035961

AR035961.1 GI:5952629

Unknown.

Unknown.

Unclassified.

1 (bases 1 to 1105)

Hastings, G. and Dillon, P.J.

Nucleic acids encoding human neuronal attachment factor-1

Patent: US 5871969-A 1 16-FEB-1999;

Location/Qualifiers

1..1105

/organism="unknown"

178 a 393 c 362 g 172 t


```
BASE COUNT      293 a   550 c   501 g   290 t
ORIGIN
Query Match      99.4%; Score 1098.6; DB 9; Length 1634;
Best Local Similarity 99.6%; Pred. No. 1.5e-149;
Matches 1101; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cgtgtctctcgtcgggtgatgaaaccccccagcccgccgccccttggtgggagagtcac 60
DQ |||||
Db 51 CGTGTCTCTCTCGCGGTGATGAAACCCAGCCGCGCGCCCTTGCGCAAGGCCCTC 110
QY 61 tgcgtctctctcgtcgggcaactctcgtcggccgcccgcgacgctcttgggggagagtcac 120
DQ |||||
Db 111 TCGGTCTCTCTCTCGCCACTCTCGCGCGCGCGCGCCAGCCCTCTTGGGGAGAGTCCATC 170
QY 121 tgttcgcagagccttgcccaaatacagcatcaccttcacgggcaagtgaagcagacg 180
DQ |||||
Db 171 TGTTCGCCAGAGCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGAGCCAGAGC 230
QY 181 gcttcccaagcagttacccctgttcgcccccttgcgagtggtcttcgtcgtcgtg 240
DQ |||||
Db 231 GCCTTCCCCAAGCAGTACCCCTGTCGCCCCCTGCGCAGTGCTTCGCTGCTGGG 290
QY 241 gcgcgcatagctcgcagctacagcgtggtggaggaaccagtagcagtaacgggctg 300
DQ |||||
Db 291 GCGGCGCATAGTCTCGACTACAGCATGTGTGGAGGAAGAACCATAGCTCAGTAACGGGCTG 350
QY 301 cgcgacttgcgagcgcgcgagccctggcgcctgctgagagagatcgagcgcgcg 360
DQ |||||
Db 351 CCGGACTTTCGGGAGCGCGCGGAGCCCTGGGGCGCTGATGAAGAGATGACAGCGCGCGGG 410
QY 361 gaggcgtcagagcgtgcgcgcgtgttctcggccgcgcgcgtcccgagcgccacggg 420
DQ |||||
Db 411 GAGGCGTGCAGAGCGTGCAGAGGTGTTTCGGCGCGCGCGCTCCCGACGCGACCGGG 470
QY 421 cagagcgtcgcgagcgtgagcgtgcagcgcagcactcgtcgtcgtctgttgggtgcgc 480
DQ |||||
Db 471 CAGAGCTCGCGGAGCTGAGAGGTGCAGCGCAGGCACTCGCTGGTCTGTTGTGTGCGC 530
QY 481 atcgtgccagcccgactggtctggtggtgagcagcctggacctctgcagcgggac 540
DQ |||||
Db 531 ATCGTGCAGCCCGCGACTGCTGCTGGCGGTGGAGACCGCTGGACCTGTGGCAGGGGAC 590
QY 541 cgttgcgggaaacagcgcgcgtgacctgtacctctacgcgcgcgcgcgcgcgcgcgcgc 600
DQ |||||
Db 591 CGTTCGGGGAAACAGCGCGCGCTGGACCTGTACCCCTACGACCGCGCGGACGACAGCGGC 650
QY 601 ttcactctctctcccaacttcgcacacttcgcacactccgcgcgcgcgcgcgcgcgcgc 660
DQ |||||
Db 651 TTCACCTTCTCTCCCGCAACTTCGCCACCATCCGCGAGGACACGGTGACCGAGATAACG 710
QY 661 tctctctctccagccacccgcaactctctctactaccgcgcgcgcgcgcgcgcgcgcgc 720
DQ |||||
Db 711 TCCTCTCTCCAGCACCCCGGCCAACTCTCTTCTACTACCGCGGCTGAAGGCCCTTGCT 770
QY 721 cccatgcgcaggggtgacactgctgcgtgcgcagagagccccagggccttctatccctccc 780
DQ |||||
Db 771 CCATCGCCAGGGTGACACTGCTGCGGTGCGACAGAGCCCGCAGGGCCTTCATCCCTCCC 830
QY 781 gccccagctctccagcagggagaaatgagattgtagacagcgcctcagttccagaaacg 840
DQ |||||
Db 831 GCCCAGGCTCTCCCGCAGGAGCAATGAGATTGTAGACAGCGGCTCAGTTCCAGAAACG 890
QY 841 ccgctggagctcgcaggtctccctgtggtcgtcgtgggagctgtgcgagcgcactgtg 900
DQ |||||
Db 891 CGCTGGACTGCGAGGTCTCCCTGTGCTGCTCTGGGACTGTGCGGAGGCCACTGTGGG 950
QY 901 aggctcgggaccaaagcagcagcactcgtacgtccggggtccagccgcgcgcgcgcgcgcgc 960
DQ |||||
Db 951 AGGCTCGGGACCAAGAGCAGGACTCGCTACGTCCGGGTCCAGCCCGCCCAACAACGCGGAGC 1010
QY 961 cccctgccccgagctcgaagaagagcgtgagtgcttccctgataactgcgtctaaagacacg 1020
```

```
Db 1011 CCCCTGCCCGAGCTCGAAGAAGAGGCTGAGTGGCTCCTCCTGATAACTCGGTCTAAGACCAG 1070
QY 1021 agccccagccccctcgtcggcccccagcagcatggggtgctcgggagcctcgtcagagct 1080
DQ |||||
Db 1071 AGCCCCAGCCCCCTGGGGCCCCCGGAGCATGGGGGTGTCGGGGGCTCCTGTGCAAGGCT 1130
QY 1081 catgctcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1105
DQ |||||
Db 1141 CATGCTGCAGGCGCGCCGAGGGCACA 1155

RESULT 3
LOCUS      AB027466      1807 bp      mRNA      PRI      05-NOV-1999
DEFINITION Homo sapiens SPON2 mRNA for spondin 2, complete cds.
ACCESSION  AB027466
VERSION     AB027466.1 GI:6172220
KEYWORDS    SPON2; spondin 2.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
REFERENCE   1 (sites)
AUTHORS     Manda,R., Kohno,T., Matsuno,Y., Takenoshita,S., Kuwano,H. and
            Yokota,J.
TITLE       Identification of genes (SPON2 and c20orf2) differentially
            expressed between cancerous and noncancerous lung cells by mRNA
            differential display
            Genomics 61 (1), 5-14 (1999)
            99443867
REFERENCE   2 (bases 1 to 1807)
AUTHORS     Yokota,J., Kohno,T. and Manda,R.
TITLE       Direct Submission
            Submitted (18-MAY-1999) to the DDBJ/EMBL/GenBank databases. Jun
            Yokota, National Cancer Center Research Institute, Biology
            Division; Tsukiji 5-chome 1-1, Chuo-ku, Tokyo 104-0045, Japan
            (E-mail: j.yokota@gan2.ncc.go.jp, Tel:81-3-3547-5272,
            Fax:81-3-3542-0807)
FEATURES    Location/Qualifiers
            source          1..1807
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="4"
            gene            276..1271
                        /map="4p16.3"
                        /gene="SPON2"
            CDS             276..1271
                        /gene="SPON2"
                        /note="alias: DIL-1; extracellular matrix protein"
                        /codon_start=1
                        /product="spondin 2"
                        /protein_id="BAA85892.1"
                        /db_xref="GI:6172221"
                        /translation="MENPSAAALGKALCALLATLGAAGQPLGGESTCSARAPAKYS
                        ITFGKMSQTAFPKOYPLRPPAQMSSLLGAHSSDYSMRKNQYVSNGLRDFAPRGE
                        ANAWLKEIEAGEALQVSAFVPSAPVSGTGTQAELEVRHSLVSVFVIVPSPD
                        WFGVDSLDCGDRWEQAALDLYPYDAGTDSGTFSSPNFATIPQDTVTITSSP
                        SHPANSFYPRLKALPIARVTLRLQSPRAFIIPAPVLPSRDNELVDSASVPETPL
                        DCEVSLWSSWGLCGGCHGLGTKSRTRYRVQVPPANNNGSPCLPEAEACVPCNV"
BASE COUNT      292 a   621 c   572 g   322 t
ORIGIN

Query Match      98.2%; Score 1084.8; DB 9; Length 1807;
Best Local Similarity 99.7%; Pred. No. 1.4e-147;
Matches 1097; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 cgtgtctcgtcgcgggtgatggaaacccccagcccgccgccccttggtgggagagtcac 60
DQ |||||
Db 258 CGTGTCTCTCTCGCGGTGATGTAAGAAACCCCGCGCGCGCCCTTGCGCAAGGCCCTC 317
QY 61 tgcgtctctcgtcgcagctctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
```


Db 1322 AACCTTTTACCTATCCAGATTGAAGACCTTCCCCCAATATGCCAAAGTCTCTTAACC 1381
 QY 745 cggctgcagacagagcccgagccttcacccctcccgccccagctcgtcccgagcaggac 804
 Db 1382 AAGATCAAGAACACACAGATCTTCAGCTTACCCATCCACCAACACAGTCCCAATCAGATC 1441
 QY 805 ---aatgagattgacagcgcctcagttccagaaacccgctggagctgcaggtctcc 861
 Db 1442 CCAAGCGGGAATGAAATAGACGGGCTCTTATAAATACCCCTTGAGACTGCCAGGTGCT 1501
 QY 862 ctgtgctcctgggagactgtgcgagagccactgtggagagctcgggagcaccagagcag 921
 Db 1502 GTGTGGTCACTTGGGGTCTTTGTAGGGTCACTGTGTGAAGAGGGGTAAACACACAGA 1561
 QY 922 actcgtcagtcctgggtccagcccgcccaaacagcgagcgcctcgcgcgagctcgaagaa 981
 Db 1562 ACCGGTATATTCACATGATCTCTGCAACACAGGTGCTCCCTGTCCATCCCTAGAGAA 1621
 982 gagcctgagtggtcctgataactgcgtctcgaagaccagagcccgagcagccc 1034
 Db 1622 AAGAGACTCTGTATCCCTGACAACTGTGTGTAATCTAAATCAGATGAGCCCTC 1674

RESULT 6
 LOCUS AR037874 400 bp DNA PAT 29-SEP-1999
 DEFINITION Sequence 1 from patent US 5804382.
 ACCESSION AR037874
 VERSION AR037874.1 GI:5956591
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 400)
 AUTHORS Sytkowski, A.J. and Yang, M.
 TITLE Methods for identifying differentially expressed genes and
 JOURNAL differences between genomic nucleic acid sequences
 PATENT: US 5804382-A 1 08-SEP-1998;
 FEATURES Location/Qualifiers
 1..400
 /organism="unknown"
 BASE COUNT 70 a 157 c 110 g 63 t
 ORIGIN
 Query Match 31.1%; Score 343.8; DB 6; Length 400;
 Best Local Similarity 97.0%; Pred. No. 1.6e-40;
 Matches 393; Conservative 0; Mismatches 7; Indels 5; Gaps 4;

QY 570 gtaccctacgacgcggagcggagcagcggttcaccttctcctcccccacacttcgcccac 629
 Db 1 GTACCCCTACGACGCGCGGAC-GACAGCGGCTTCACCTTCTCTCCGCCCAACTTCGCCAC 59
 QY 630 catcccgaggacacggtgacagagataacgtctcctcctcctccagcaccacccgccaactc 689
 Db 60 CATCCCGCA-GACACGGTGACCGAGATAACGTCTCTCTCCAGCCACCGCGCAGGTC 118
 QY 690 ctctactaccgcgcgtggaagccctgcctccatccagcaggtgacactggtcggct 749
 Db 119 CTCTACTACCGCGGCTGAAGGCC--TGCTCCCATCCAGGCTGACACTGGTGGGCT 176
 QY 750 gcagacagcccgagcgttcacccctcccgcccgagctcctccagcagcaggaacaatga 809
 Db 177 GCGACAGAGCCCGCAGGCGCTTCATCTCTCCGCCCGCAGTCTGCCAGCA-GGACAAATGA 235
 QY 810 gatttagacagcgcctcagttccagaaacccgctggagctgcaggtctcctctgtggtc 869
 Db 236 GATTGTAGACAGCGCTCAGTTCAGAAACCGCCCTGGACTGCGAGGCTCTCCCTGTGGTC 295
 QY 870 gtcctgggagctgtgcgagcactgtggagctcgggagcaccagagcagactcgcta 929
 Db 296 GTCTTGGGAGCTGTGGAGGCCACTGTGTGTAATCTAAATCAGATGAGCCCTC 355

QY 930 cgtccgggtccagcccgcaaacacggagcggagcctcctgccccagct 974
 Db 356 CGCCCGGTCCAGCCCGCCCAACACGGGAGCCCTGCCCCGAGCT 400
 RESULT 7
 LOCUS AR125230 400 bp DNA PAT 16-MAY-2001
 DEFINITION Sequence 1 from patent US 6177244.
 ACCESSION AR125230
 VERSION AR125230.1 GI:14111292
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 400)
 AUTHORS Sytkowski, A.J. and Yang, M.
 TITLE NFG-1 gene that is differentially expressed in prostate tumors
 JOURNAL Patent: US 6177244-A 1 23-JAN-2001;
 FEATURES Location/Qualifiers
 1..400
 /organism="unknown"
 BASE COUNT 69 a 159 c 109 g 63 t
 ORIGIN

Query Match 30.7%; Score 339; DB 6; Length 400;
 Best Local Similarity 96.3%; Pred. No. 8e-40;
 Matches 390; Conservative 0; Mismatches 10; Indels 5; Gaps 4;
 QY 570 gtaccctacgacgcggagcggagcagcggttcaccttctcctcccccacacttcgcccac 629
 Db 1 GTACCCCTACGACGCGCGGAC-GACAGCGGCTTCACCTTCTCTCCGCCCAACTTCGCCAC 59
 QY 630 catcccgaggacacggtgacagagataacgtctcctcctcctccagcaccacccgccaactc 689
 Db 60 CATCCCGCA-GACACGGTGACCGAGATAACGTCTCTCTCCAGCCACCGCGCAGGTC 118
 QY 690 ctctactaccgcgcgtggaagccctgcctccatccagcaggtgacactggtcggct 749
 Db 119 CTCTACTACCGCGGCTGAAGGCC--TGCTCCCATCCAGGCTGACACTGGTGGGCT 176
 QY 750 gcagacagcccgagcgttcacccctcccgcccgagctcctccagcagcaggaacaatga 809
 Db 177 GCGACAGAGCCCGCAGGCGCTTCATCTCTCCGCCCGCAGTCTGCCAGCA-GGACAAATGC 235
 QY 810 gatttagacagcgcctcagttccagaaacccgctggagctgcaggtctcctctgtggtc 869
 Db 236 GATTGTAGACAGCGCTCAGTTCAGAAACACCGCTGGACTGCGAGGCTCTCCCTGTGGTC 295
 QY 870 gtcctgggagctgtgcgagcactgtggagctcgggagcaccagagcagactcgcta 929
 Db 296 GTCTTGGGAGCTGTGGAGGCCACTGTGTGTAATCTAAATCAGATGAGCCCTC 355
 QY 930 cgtccgggtccagcccgcccaaacacggagcggagcctcctgccccagct 974
 Db 356 CGCCCGGTCCAGCCCGCCCAACACGGGAGCCCTGCCCCGAGCT 400

RESULT 8
 LOCUS AC022763/c 197176 bp DNA HTG 27-APR-2000
 DEFINITION Homo sapiens clone RP11-296G16, WORKING DRAFT SEQUENCE, 24
 unordered pieces.
 ACCESSION AC022763
 VERSION AC022763.2 GI:7652738
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 197176)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone Rp11-296G16
Unpublished
2 (bases 1 to 197176)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelra,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Melidrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 27, 2000 this sequence version replaced gi:6987625.
All repeats were identified using RepeatMasker:
Sait, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5843
Center clone name: 296_G_16
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 185565 bases at least Q40
Consensus quality: 190999 bases at least Q30
Consensus quality: 192979 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 194876; sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 438: contig of 438 bp in length
* 439 538: gap of 100 bp
* 539 2494: contig of 1956 bp in length
* 2495 2594: gap of 100 bp
* 2595 3740: contig of 1146 bp in length
* 3741 3840: gap of 100 bp
* 3841 6181: contig of 2341 bp in length
* 6182 6281: gap of 100 bp
* 6282 8921: contig of 2640 bp in length
* 8922 9021: gap of 100 bp
* 9022 12132: contig of 3111 bp in length
* 12133 12232: gap of 100 bp
* 12233 15557: contig of 3325 bp in length
* 15558 15657: gap of 100 bp
* 15658 19986: contig of 4329 bp in length
* 19987 20086: gap of 100 bp
* 20087 21983: contig of 1897 bp in length

FEATURES
source
1..197176
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rp11-296G16"
/clone.lib="RPC1-11 Human Male BAC"
1..438
/note="assembly_fragment"
clone_end:T7
vector_side:left
539..2494
/note="assembly_fragment"
2595..3740
/note="assembly_fragment"
3841..6181
/note="assembly_fragment"
6282..8921
/note="assembly_fragment"
9022..12132
/note="assembly_fragment"
12233..15557
/note="assembly_fragment"
15658..19986
/note="assembly_fragment"
20087..21983
/note="assembly_fragment"
clone_end:SP6
vector_side:left
22084..26850
/note="assembly_fragment"
26951..31317
/note="assembly_fragment"
31418..36013
/note="assembly_fragment"
36114..41819
/note="assembly_fragment"
41920..46044
/note="assembly_fragment"
46145..52849
/note="assembly_fragment"
52950..63466
/note="assembly_fragment"
63567..70466
21984 22083: gap of 100 bp
22084 26850: contig of 4767 bp in length
26851 31317: contig of 4367 bp in length
31318 31417: gap of 100 bp
31418 36013: contig of 4596 bp in length
36014 36113: gap of 100 bp
36114 41819: contig of 5706 bp in length
41820 41919: gap of 100 bp
41920 46044: contig of 4125 bp in length
46045 46144: gap of 100 bp
46145 52849: contig of 6705 bp in length
52850 52949: gap of 100 bp
52950 63466: contig of 10517 bp in length
63467 63566: gap of 100 bp
63567 70466: contig of 6900 bp in length
70467 70566: gap of 100 bp
70567 80411: contig of 9845 bp in length
80412 80511: gap of 100 bp
80512 92230: contig of 11719 bp in length
92231 92330: gap of 100 bp
92331 105063: contig of 12733 bp in length
105064 105163: gap of 100 bp
105164 115199: contig of 10036 bp in length
115200 115299: gap of 100 bp
115300 127901: contig of 12602 bp in length
127902 128001: gap of 100 bp
128002 145911: contig of 17910 bp in length
145912 146011: gap of 100 bp
146012 197176: contig of 51165 bp in length.

Kazusa DNA Research Institute.
 FEATURES
 source Location/Qualifiers
 1. .4292
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="as00108"
 /dev_stage="adult"
 /tissue_type="spleen"
 /note="vector:pluascriptii SK plus"
 1. .653
 /gene="FLJ00108"
 <1. .653
 /gene="FLJ00108"
 /note="For this clone, GeneMark analysis triggered an alert for spurious CDS split. The result of GeneMark analysis of the nucleotide sequence of this clone is accessible through http://www.kazusa.or.jp/NEDO. Start codon is not identified."
 /codon_start=3
 /product="FLJ00108 protein"
 /protein_id="BAB15789.1"
 /db_xref="GI:10440518"
 /translation="SAELEVRHSLVSFVVRVPDWFVGVDSLDLDCDGRWRQEA ALDLYPDAGTDSGFTFSSPNFATIFQDVTVEYSGHPACARAMAPAVGTGLPGCTL ASQPRTAACGFCVEALLLAPAVSKGVQPSHTLPTRGHSTAGTAGRASGRGPFILAS VSSPCSVGMLRGVHGTLSPTVQVRWTLGSGPSEAIRIPEGGLLGPQANLFLPR"
 BASE COUNT 788 a 1323 c 1349 g 832 t
 ORIGIN
 Query Match 25.2%; Score 279; DB 9; Length 4292;
 Best Local Similarity 100.0%; Pred. No. 1.1e-31;
 Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 827 cagttccagaaccccgctgactgcaggtctccctgtgtctctctggtgactgtgcg 886
 Db 3568 CAGTTCAGAAACCGCTGGACTCGGAGTCTCCCTGTGCTGCTCTGGGACTGTGCG 3627
 Qy 887 gagccactgtgagactcggagaccagagagactcgtactcgcggtccagcccg 946
 Db 3628 GAGGCCACTGTGGAGGCTCGGGACCAAGACAGAGACTGCTAGTCTCGGGTCCAGCCCG 3687
 Qy 947 ccaacaacggagccctgcccgagctcgaagaagaggtgagtgctcctgataact 1006
 Db 3688 CCAACAACGGAGCCCTGCCCGAGCTCGAAGAAGAGGCTGAGTGGCTCCCTGATAACT 3747
 Qy 1007 gcgtctaagaccagagcccccagagccctggggcccccccgagcagcatgggtgcggggg 1066
 Db 3748 GCGTCTAAGACCAGAGCCCGCAGCCCTGGGGCCCCCGGAGCCATGGGGGTGTCGGGGG 3807
 1067 ctctgtcaggtcactatgtcagcggccgagggcaca 1105
 Db 3808 CTCCTGTGAGGCTCATGCTGCAGGCGCCGAGGGCACA 3846
 RESULT 11
 AK026054
 LOCUS 2820 bp mRNA PRI 29-SEP-2000
 DEFINITION Homo sapiens CDNA: FLJ22401 fls, clone HRC08032, highly similar to AB027466 Homo sapiens SPON2 mRNA for spondin 2.
 AK026054
 ACCESSION AK026054.1 GI:10438772
 VERSION
 KEYWORDS oligo capping; fls (full insert sequence).
 SOURCE Homo sapiens primary human renal epithelial cells cDNA to mRNA, clone_lib:HRC clone:HRC08032.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (sites)
 Kawanaba A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T. and Sugano S.
 NEDO human cDNA sequencing project

Unpublished (2000)
 2 (bases 1 to 2820)
 SUGANO S., SUZUKI Y., OTA T., OBAAYASHI M., NISHI T., ISOGAI T., SHIBAHARA T., TANAKA T. and NAKAMURA Y.
 Direct Submission
 Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
 NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
 FEATURES
 Location/Qualifiers
 1. .2820
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="primary human renal epithelial cells"
 /clone="HRC08032"
 /clone_lib="HRC"
 /note="cloning vector pME18SFL3"
 1. .2820
 /note="highly similar to AB027466 Homo sapiens SPON2 mRNA for spondin 2"
 BASE COUNT 535 a 849 c 873 g 563 t
 ORIGIN
 Query Match 25.1%; Score 277.4; DB 9; Length 2820;
 Best Local Similarity 99.6%; Pred. No. 2.3e-31;
 Matches 278; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 827 cagttccagaaccccgctgactgcaggtctccctgtgtgctcctggtgactgtgcg 886
 Db 2079 CAGTTCAGAAACCGCTGGACTCGGAGTCTCCCTGTGCTGCTCTGGGACTGTGCG 2138
 Qy 887 gagccactgtgagactcggagaccagagagactcgtactcgcggtccagcccg 946
 Db 2139 GAGGCCACTGTGGAGGCTCGGGACCAAGACAGAGACTGCTAGTCTCGGGTCCAGCCCG 2198
 Qy 947 ccaacaacggagccctgcccgagctcgaagaagaggtgagtgctcctgataact 1006
 Db 2199 CCAACAACGGAGCCCTGCCCGAGCTCGAAGAAGAGGCTGAGTGGCTCCCTGATAACT 2258
 Qy 1007 gcgtctaagaccagagcccccagagccctggggcccccccgagcagcatgggtgtcggggg 1066
 Db 2259 GCGTCTAAGACCAGAGCCCGCAGCCCTGGGGCCCCCGGAGCCATGGGGGTGTCGGGGG 2318
 Qy 1067 ctctgtcaggtcactatgtcagcggccgagggcaca 1105
 Db 2319 CTCCTGTGAGGCTCATGCTGCAGGCGCCGAGGGCACA 2357
 RESULT 12
 AB006084
 LOCUS 3381 bp mRNA VRT 10-MAY-2001
 DEFINITION Danio rerio mRNA for MINDIN1, complete cds.
 ACCESSION AB006084
 VERSION AB006084.1 GI:2529220
 KEYWORDS MINDIN1.
 SOURCE Danio rerio embryo cDNA to mRNA.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.
 1 (sites)
 Higashijima S., Nose A., Eguchi G., Hotta Y. and Okamoto H.
 REFERENCES
 AUTHORS
 TITLE Mindin/F-spondin family: novel ECM proteins expressed in the

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

zebrafish embryonic axis
Dev. Biol. 192 (2), 211-227 (1997)
98104230
2 (bases 1 to 381)
Higashijima, S.
Direct Submission
Submitted (28-JUL-1997) Shin-ichi Higashijima, National Institute
for Basic Biology, Division of Morphogenesis; Nishigonaka38,
Myodaijicho, Okazaki, Aichi 444, Japan (E-mail: shinichi@nibb.ac.jp,
Tel:81-564-55-7372, Fax:81-564-55-7571)

```

162.161 304 337.3727, rxn:01.304.337.3751J)
Location/Qualifiers
1. .3381
/organism="Danio rerio"
/db_xref="taxon:7955"
/dev_stage="embryo"
48. .1052
CDS
/codon_start=1
/product="MIND1"
/protein_id="BAA22808.1"
/db_xref="GI:252921"
/translation="MSSSILVPGWLQQLVLLRFTLSCAALVNSTNGTECSARGPA
SYIVFTGHKSPOTFFKQYPLFRPAQSKLVMTVHNEQYRLWEGAPASDGKMSFAE
QGLTVLDVKAEKARRSRVSGVMTAGTSPGIGHSTSEVLLTPRSLPVLIVKLTPS
PDWFVGDGVLNCEGKQKQEVTFDLHFDAGTGSFTFSSNPPTLPPTPNITWTSQ
KPNHPSANSTYYRLNELPLATVWKRQSRPLVRQGNRLSNHLPDASKPHRTSETPL
DCEYPMSSWGGLCFGCARGGLRHRTYILLKPNASGSPCELEOEECTPHNCLADO

```

```

polyA_site      3381
                 /note="11 a nucleotides"
BASE COUNT      949 a      713 c      707 g      1012 t
ORIGIN

```

Query Match 23.3%; Score 258; DB 5; Length 3381;
Best Local Similarity 57.1%; Pred. No. 1.3e-28;
Matches 512; Conservative 0; Mismatches 375; Indels 9; Gaps 2;

127	QY	gccagagccctggccaaatacacagatacaacctcaacgttcaacgggccaagtggagccagacggccttc	186
128			
129			
130			
131			
132			
133			
134			
135			
136			
137			
138			
139			
140			
141			
142			
143			
144			
145			
146			
147			
148			
149			
150			
151			
152			
153			
154			
155			
156			
157			
158			
159			
160			
161			
162			
163			
164			
165	DB	GCTCAGAGACCTGCTATCTTACATCGTGGTGTTCACCGGACACTGAGTCCACAGACACTTC	224
166			
167			
168			
169			
170			
171			
172			
173			
174			
175			
176			
177			
178			
179			
180			
181			
182			
183			
184			
185			
186			
187	QY	cccaagcagtagccccctgtctccgcccccctggccagtggtctctcgtctggtgggcccgcg	246
188			
189			
190			
191			
192			
193			
194			
195			
196			
197			
198			
199			
200			
201			
202			
203			
204			
205			
206			
207			
208			
209			
210			
211			
212			
213			
214			
215			
216			
217			
218			
219			
220	DB	CCTAAACAGTATCCCTTATTCGGGCCACCTGCCAGTGGTCCAAACTCATGGTTGTAACC	284
221			
222			
223			
224			
225			
226			
227			
228			
229			
230			
231			
232			
233			
234			
235			
236			
237			
238			
239			
240			
241			
242			
243			
244			
245			
246			
247	QY	catagctccgactacagcatgtggaggaagaacacagtcagtcagtaacgggctgcgcgac	306
248			
249			
250			
251			
252			
253			
254			
255			
256			
257			
258			
259			
260			
261			

Qy	667	ttctccagccaccgcgccaactcttctactaccgcggtgaagccctgctcccatc	726
Db	702	AAGCAAAACCAACCCGGGAAATTCCTTTTATTATCTCGGTAATAACTGCCACCTCT	760
Qy	727	gccagggtgacactggctgcgctgcacacagcccgaggccttcacccctccgcgccca	786
Db	761	-----GCCAACCATATGGGTGAACGACATCCCGCTTGCTCTCCGTCAACAGAACCGG	815
Qy	787	gtctgccccagcagggaacaatgagattgtagacagcgctcagttccagaaacgcgcgtg	846
Db	816	CTGTCCAATCACATCTTACCAGACGCAAGCAAAACCTCACAGGTTTTCAGAAACGCGCTG	875
Qy	847	gactgcagatctccctgtgctcctgggagactatgcgagccactgtggagagctc	906
Db	876	GACTGCGAGGTGTCGATGTGTCCTCTGGGGTCTTGTTTTGGCCCGTCGCTCGAGGC	935
Qy	907	gggaccagagcaggaactgctcagtcgcgggtccagcccgcccaacacgaggagccctgc	966
Db	936	GGCTACGCCATCGCATCTGTTACATCTGCTGAAACCTGCCAACAGCGTTGCGCCATGT	995
Qy	967	cccagactcaaaagaggctgagtcgctccctgataaactgcgtctaaagacagag	1022
Db	996	CCCGAGCTGGAGGAACGAGGAGGATGTCACCCGCAACTGCCTGCAGATCATGT	1051
RESULT 13			
LOCUS	AR035967	316 bp	DNA
DEFINITION	Sequence 16 from patent US 5871969.		PAT
ACCESSION	AR035967		
VERSION	AR035967.1	GI:5952635	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 316)		
AUTHORS	Hastings,G; and Dillon,P.J.		
TITLE	Nucleic acids encoding human neuronal attachment factor-1		
JOURNAL	Patent: US 5871969-A 16 16-FEB-1999;		
FEATURES	Location/Qualifiers		
source	1..316		
	/organism="unknown"		
BASE COUNT	48 a	84 c	120 g
ORIGIN	49 t		15 others

Query Match	18.3%;	Score 202.6;	DB 6;	Length 316;
Best Local Similarity	85.6%;	Pred. No. 4e-20;		
Matches 274;	Conservative 0;	Mismatches 39;	Indels 7;	Gaps
QY 272	ggaagaacacagtcagtcagtaaacgggctgcgcgaactttgcgagacgcgcgagggcctggg	331		
Db 1	GGCANNGCCAGTAGCTCA-TAACGGGCTGCGGACTTTCGGGA-NGCGCGAGGCTGGG	58		
QY 332	cgcctgatgaaggagatcgaggcggcggggagggcgctgcagagcgtgcacgcgctgttt	391		
Db 59	CGCTGATGAGGAGATCAGGCGCGCGGGAGGCGCTGCAGAG-GTGCACGAGGTGTTTT	117		
QY 392	cggcgccgcgcgtccccagcggcacgcgggcagacgctcggcgagctgtaggtgcagcgca	451		
Db 118	CGCGCGCGG-GTNNCCAGCGNCACNCGCAGACGTGCGCGAACTGGNAGGTGCAGCGCA	176		
QY 452	ggcactcgctggtctogtttgtgtgcgcatacgtgccacgcccgactggttcgtggcgcg	511		
Db 177	GGCACATCGCTGTGTCTGTTTGTGTGTGCGCATCTGCACCAGCCCGACTGTTTCGTGGGCG	236		
QY 512	tggacagcctggacc---tgtgcgacgggggacgcgttggcggaacagcgcgcgctggacc	568		
Db 237	TGGACAGCCTGGGACCTGTGANNACGGGAGCACTTTNGCNGNNAACAGCGCNGCTTGGACC	296		
QY 569	tgtaccacctacagcgcgcggg	588		
Db 297	TGTANCCCTACGACGNCGGG	316		

[illegible]

Tue Dec 11 09:43:24 2001

us-09-170-042a-1.rge

Page 12

Search completed: December 10, 2001, 19:46:49
Job time: 22163 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2001, 13:37:26 ; Search time 164.06 Seconds
(without alignments)
5774.378 Million cell updates/sec

Title: US-09-170-042A-1
Perfect score: 1105
Sequence: 1 cgcgtctctgcgggtgat.....tcgagcgccgagggcaca 1105

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
All number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1105	100.0	1105	20	AA02019 Human NAF-1 DNA.
2	1100.2	99.6	1105	18	AA078360 Human neuronal att
3	1100.2	99.6	1840	21	AA0295034 Cancer specific ge
4	1100	99.5	1848	22	AA034981 Human colon cancer
5	1098.6	99.4	1607	19	AA063241 CDNA encoding an a
6	1098.6	99.4	1770	22	AA090567 Human extracellular
7	1098.6	99.4	1785	22	AA090566 Human extracellular
8	1098.6	99.4	1846	20	AA034089 Human PRO866 nucle
9	1098.6	99.4	1847	21	AA078507 Human PRO866 (UN04
10	1098.6	99.4	1847	21	AA058630 Human PRO866 prote
11	1098.6	99.4	1847	21	AAA49728 Human PRO866 cDNA

C	12	1098.6	99.4	1926	22	AA159814	Human polynucleoti
	13	1098.6	99.4	1991	22	AA158028	Human polynucleoti
	14	1088.2	98.5	1718	21	AA059794	Human secreted pro
	15	989.6	89.6	996	20	AA072537	Human mindin cDNA.
	16	843.8	76.4	1458	21	AA025275	Human secreted pro
	17	681	61.6	993	19	AA063258	Degenerate sequence
	18	341.8	31.1	400	19	AA053726	Nucleotide sequenc
	19	343.6	30.9	1021	20	AA072538	Human mindin-relat
	20	339	30.7	400	20	AA060948	Human NPG-1 partia
	21	300	27.1	506	20	AA02024	Human NAF-1 DNA ho
	22	279	25.2	2261	21	AA059258	Human secreted pro
	23	233.4	21.1	432	19	AA063270	EST sequence encod
	24	204	18.5	204	21	AA095016	Prostate cancer sp
	25	202.6	18.3	316	20	AA02026	Human NAF-1 DNA ho
	26	202.6	18.3	316	20	AA02025	Human NAF-1 DNA ho
	27	198.4	18.0	553	19	AA063268	EST sequence encod
	28	176.8	16.0	539	22	AA157550	Human colorectal c
	29	173.2	15.7	541	19	AA063266	EST sequence encod
	30	142.2	12.9	470	19	AA063267	EST sequence encod
	31	78	7.1	4061	20	AA030094	Bovine vascular sm
	32	71.6	6.5	3079	20	AA030093	Human vascular smo
	33	71.6	6.5	3079	22	AA094993	Human ovarian canc
	34	71.6	6.5	3999	22	AA094986	Human ovarian canc
	35	70	6.3	4029	14	AA052674	F-spondin (FP5-9)
	36	70	6.3	4029	20	AA030095	Rat vascular smoot
	37	61.6	5.6	3226	14	AA052675	F-spondin coding s
	38	61.2	5.5	114955	20	AA053491	Human adenosine A1
	39	57.2	5.2	1816	14	AA052676	Partial sequence o
	40	57.2	5.2	1848	22	AA034981	Human colon cancer
	41	57	5.2	1105	18	AA078360	Human neuronal att
	42	57	5.2	1105	20	AA02019	Human NAF-1 DNA.
	43	57	5.2	1458	21	AA052575	Human secreted pro
	44	57	5.2	1718	21	AA059794	Human secreted pro
	45	57	5.2	1926	22	AA159814	Human polynucleoti

ALIGNMENTS

RESULT 1
AA02019
ID AA02019 standard; DNA; 1105 BP.
XX
AC AA02019;
XX
DT 21-APR-1999 (first entry)
XX
DE Human NAF-1 DNA.

XX NAF-1; neuronal attachment factor-1; F-spondin analogue; treatment;
KW spinal cord injury; peripheral nerves damage; neural cell adhesion;
KW neurite extension; tumour cell metastasis; inhibitor; mobility; disease;
KW endothelial cell proliferation; tumour neovascularisation; haemostasis;
KW angiostatic agent; wound healing; diagnostic; neurotrophic; anticancer;
KW antimetastatic; anti-angiogenic; antimalarial; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
CDS 19..1014
FT /*tag= a
FT /*product= "NAF-1"
FT 19..87
FT /*tag= b
FT 88..1011
FT /*tag= c

XX US5871969-A.

XX 16-FEB-1999.

XX 12-FEB-1997; 97US-0799173.

XX

CC polypeptides, polynucleotides encoding them (especially encoding
 CC amino acids 1-331, 21-331 or 27-331 of RGL), methods for
 CC producing the polypeptides, expression vectors and genetically
 CC engineered host cells for expression of the polypeptides. It
 CC also provides methods for utilizing the polynucleotides and
 CC polypeptides in research, diagnosis and therapeutic applications.
 CC Thus, RGL polynucleotides can be used in the production of
 CC recombinant proteins, in methods for assessing RGL expression
 CC in cells, and in the development of ribozyme and antisense
 CC oligonucleotides useful for administration e.g. to a human
 CC patient for treatment of a disease such as prostate cancer or
 CC benign prostatic hyperplasia, which is alleviated by decreasing
 CC the level of RGL activity.

XX
 SQ Sequence 1770 BP; 285 A; 614 C; 570 G; 301 T; 0 other;

Query Match 99.4%; Score 1098.6; DB 22; Length 1770;
 Best Local Similarity 99.6%; Pred. No. 5.3e-196;
 Matches 1101; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cgtctctctcgtcgtgtgatgaaacccacgcccgcgcgcgccttggtgcaagccctc 60
 |||||
 Db 278 cgtctctcgtcgtgtgatgaaacccacgcccgcgcgcgccttggtgcaagccctc 337
 |||||
 QY 61 tgcgtctctctcgtcgtcactctggcgccgcgcgcgcctcttgggggagatccatc 120
 |||||
 Db 338 tgcgtctctcgtcgtcactctggcgccgcgcgcgcctcttgggggagatccatc 397
 |||||
 QY 121 tttccgcagagccctggccaaatcacagatcaccttcacagggcgaagtgcagacag 180
 |||||
 Db -398 tgttcgcgcagagcccgccaaatcacagatcaccttcacagggcgaagtgcagacag 457
 |||||
 QY 181 gcttccccaagcagctacccctgttcgcgcgcgcgcctgcgcgcctcttgcgtggtg 240
 |||||
 Db 458 gcttccccaagcagctacccctgttcgcgcgcgcgcctgcgcgcctcttgcgtggtg 517
 |||||
 QY 241 gccgcgcatagctcgcgtacagatgtggaggagaaaccagtcagtaacgggctg 300
 |||||
 Db 518 gccgcgcatagctcgcgtacagatgtggaggagaaaccagtcagtaacgggctg 577
 |||||
 QY 301 cgcagcttcgcagagcgcgcgcgcgccttcgcgcctgatgaaggagatcagagcgcgcgg 360
 |||||
 Db 578 cgcgacttcgcagagcgcgcgcgcgccttcgcgcctgatgaaggagatcagagcgcgcgg 637
 |||||
 QY 361 gaggcgctgcagagcgtcagcgcgcgtgttctgcgcgcgcgccttcgcgcgcgcgcgcgc 420
 |||||
 Db 638 gaggcgctgcagagcgtcagcgcgcgtgttctgcgcgcgcgccttcgcgcgcgcgcgcgc 697
 |||||
 QY 421 cagacgtcgc 480
 |||||
 Db 698 cagacgtcgc 757
 |||||
 QY 481 atcgtgccagcccgacgttgcgtggcggtgacagcctggacctgacagcgcgcgcgcgcgc 540
 |||||
 Db 758 atcgtgccagcccgacgttgcgtggcggtgacagcctggacctgacagcgcgcgcgcgcgc 817
 |||||
 QY 541 cgttgc 600
 |||||
 Db 818 cgttgc 877
 |||||
 QY 601 ttacacttct 660
 |||||
 Db 878 ttacacttct 937
 |||||
 QY 661 tct 720
 |||||
 Db 938 tct 997
 |||||
 QY 721 cccatcgc 780
 |||||
 Db 998 cccatcgc 1057
 |||||

QY 781 gccccagtctcgtccacgagcaggacaatgagattgtagacagcctcagttccagaacg 840
 |||||
 Db 1058 gccccagtctcgtccacgagcaggacaatgagattgtagacagcctcagttccagaacg 1117
 |||||
 QY 841 ccgctgacgtcgcaggtctccctctggtctcgtctggtgagactgtgcgagggcactgtggg 900
 |||||
 Db 1118 ccgctgacgtcgcaggtctccctctggtctcgtctggtgagactgtgcgagggcactgtggg 1177
 |||||
 QY 901 aggtcgggaccagagcagcagactcgtactcgtcgggtccagcccccacacacgggagc 960
 |||||
 Db 1178 aggtcgggaccagagcagcagactcgtactcgtcgggtccagcccccacacacgggagc 1237
 |||||
 QY 961 cctctccccagctcgaagaagaggctgagtgcgtccctcgataactgcgtctaagaccag 1020
 |||||
 Db 1238 cctctccccagctcgaagaagaggctgagtgcgtccctcgataactgcgtctaagaccag 1297
 |||||
 QY 1021 agcccgacgacctggggcccccagcagcagcagcagcagcagcagcagcagcagcagcagc 1080
 |||||
 Db 1298 agcccgacgacctggggcccccagcagcagcagcagcagcagcagcagcagcagcagcagc 1357
 |||||
 QY 1081 catgctcagc 1105
 |||||
 Db 1358 catgctcagc 1382
 |||||

RESULT 7
 AAF90566
 ID AAF90566 standard; DNA; 1785 BP.
 XX
 AC AAF90566;
 XX
 DT 22-AUG-2001 (first entry)
 XX
 DE Human extracellular matrix protein RGL DNA.
 XX
 KW RGL; human; extracellular matrix protein; prostate cancer;
 KW metastasis; tumour; benign prostatic hyperplasia; gene therapy;
 KW diagnosis; antitumour; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 296..1294
 FT /*tag= a
 XX
 PN WO200144291-A2.
 XX
 PD 21-JUN-2001.
 XX
 PF 15-DEC-2000; 2000WO-US33901.
 XX
 PR 16-DEC-1999; 99US-0172370.
 PR 07-DEC-2000; 2000US-9966561.
 XX
 PA (SCHD) SCHERING AG.
 XX
 PI Har'ins R, Parkes D, Parry G, Schneider DW, Steinbrecher R;
 CR WPI: 2001-398128/42.
 DR P-PDB; AAB82472.
 XX
 PT Novel human extracellular matrix polypeptide, RGL, useful in research,
 PT diagnosis and treatment of metastasis such as prostate cancer -
 XX
 PS Claim 8; Fig 1; 69pp; English.
 XX
 CC The present sequence is that of a polynucleotide, designated RGL,
 CC encoding novel human extracellular matrix polypeptide RGL (see
 CC AAB82472). The full-length coding sequence was deduced from a
 CC clone identified from a database screening that was found in the
 CC category of cell adhesion molecules and described as a homologue
 CC of f-spondin. RGL is expressed in prostate tissue and may be
 CC over-expressed in prostate tumours. It shows homology to a

CC superfamily of extracellular matrix proteins encoded by the
 CC mindin/f-spondin genes. The invention provides human Rgl
 CC polypeptides, polynucleotides encoding them (especially encoding
 CC amino acids 1-331, 21-331 or 27-331 of Rgl), methods for
 CC producing the polypeptides, expression vectors and genetically
 CC engineered host cells for expression of the polypeptides. It
 CC also provides methods for utilizing the polynucleotides and
 CC polypeptides in research, diagnosis and therapeutic applications.
 CC Thus, rgl polynucleotides can be used in the production of
 CC recombinant proteins, in methods for assessing Rgl expression
 CC in cells, and in the development of ribozyme and antisense
 CC oligonucleotides useful for administration e.g. to a human
 CC patient for treatment of a disease such as prostate cancer or
 CC benign prostatic hyperplasia, which is alleviated by decreasing
 CC the level of Rgl activity.

XX Sequence 1785 BP; 300 A; 614 C; 570 G; 301 T; 0 other;

Query Match 99.4%; Score 1098.6; DB 22; Length 1785;
 Best Local Similarity 99.6%; Pred. No. 5.3e-196;
 Matches 1101; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cgtctctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 60
 Db 278 cgtctctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 337
 Qy 61 tgcgtctctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 120
 Db 338 tgcgtctctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 397
 Qy 121 tgcgtctctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 180
 Db 398 tgcgtctctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 457
 Qy 181 gctctctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 240
 Db 458 gctctctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 517
 Qy 241 gccgctctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 300
 Db 518 gccgctctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 577
 Qy 301 cgcgactctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 360
 Db 578 cgcgactctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 637
 Qy 361 gaggcgtctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 420
 Db 638 gaggcgtctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 697
 Qy 421 cagacgtctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 480
 Db 698 cagacgtctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 757
 Qy 481 atcgtctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 540
 Db 758 atcgtctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 817
 Qy 541 cgttgcgggaacagcggcgtgatggaaccccccagcccgcccgccctggcgaagccctc 600
 Db 818 cgttgcgggaacagcggcgtgatggaaccccccagcccgcccgccctggcgaagccctc 877
 Qy 601 ttcacct 660
 Db 878 ttcacct 937
 Qy 661 tct 720
 Db 938 tct 997
 Qy 721 cccatcgcagggtgacacttggtgcggctgcgacagagcccccgggcttcatccctccc 780

Db 998 cccatcgcagggtgacacttggtgcggctgcgacagagccccagggccttcatccctccc 1057
 Qy 781 gccacgtct 840
 Db 1058 gccacgtct 1117
 Qy 841 ccgctggactcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 900
 Db 1118 ccgctggactcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 1177
 Qy 901 agctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 960
 Db 1178 agctctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 1237
 Qy 961 cccctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 1020
 Db 1238 cccctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 1297
 Qy 1021 agcccgagccccctggggcccccccgagagccatgggtgtcgggggctcctgtgcaggct 1080
 Db 1298 agcccgagccccctggggcccccccgagagccatgggtgtcgggggctcctgtgcaggct 1357
 Qy 1081 catgctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 1105
 Db 1358 catgctcgcgggtgatggaaccccccagcccgcccgccctggcgaagccctc 1382
 RESULT 8
 AAZ34089
 ID AAZ34089 standard; cDNA; 1846 BP.
 XX AAZ34089;
 AC AAZ34089;
 DT 07-DEC-1999 (first entry)
 XX Human PRO866 nucleotide sequence.
 DE Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
 KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
 KW secreted protein; transmembrane protein; ss.
 OS Homo sapiens.
 PN WO9946281-A2.
 XX 16-SEP-1999.
 XX 08-MAR-1999; 99WO-US05028.
 PR 10-MAR-1998; 98US-0077450.
 PR 11-MAR-1998; 98US-0077632.
 PR 11-MAR-1998; 98US-0077641.
 PR 11-MAR-1998; 98US-0077649.
 PR 12-MAR-1998; 98US-0077791.
 PR 13-MAR-1998; 98US-0078004.
 PR 17-MAR-1998; 98US-0040220.
 PR 20-MAR-1998; 98US-0078886.
 PR 20-MAR-1998; 98US-0078910.
 PR 20-MAR-1998; 98US-0078936.
 PR 20-MAR-1998; 98US-0078939.
 PR 25-MAR-1998; 98US-0079294.
 PR 26-MAR-1998; 98US-0079656.
 PR 27-MAR-1998; 98US-0079663.
 PR 27-MAR-1998; 98US-0079664.
 PR 27-MAR-1998; 98US-0079689.
 PR 27-MAR-1998; 98US-0079728.
 PR 27-MAR-1998; 98US-0079786.
 PR 30-MAR-1998; 98US-0079920.
 PR 30-MAR-1998; 98US-0079923.
 PR 31-MAR-1998; 98US-0080105.
 PR 31-MAR-1998; 98US-0080165.
 PR 31-MAR-1998; 98US-0080194.

0:

them. The tumour is especially a ca-

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2001, 13:37:26 ; Search time 89.78 Seconds
(without alignments)
2787.460 Million cell updates/sec

Title: US-09-170-042A-1
Perfect score: 1105
Sequence: 1 cgcgtctctgcgggtgat.....tgcaggcgccgagggcaca 1105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1105	100.0	1105	2	US-08-799-173A-1
2	1051.8	95.2	1779	4	US-09-371-696-1
3	343.8	31.1	400	1	US-08-684-326-1
4	339	30.7	400	4	US-09-022-238-1
5	300	27.1	506	2	US-08-799-173A-15
6	202.6	18.3	316	2	US-08-799-173A-16
7	202.6	18.3	316	2	US-08-799-173A-17
8	70	6.3	4029	1	US-07-862-021B-9
9	70	6.3	4029	1	US-08-313-288B-9
10	70	6.3	4029	5	PCT-US93-03164-9
11	61.6	5.6	3226	1	US-07-862-021B-11
12	61.6	5.6	3226	1	US-08-313-288B-11
13	61.6	5.6	3226	5	PCT-US93-03164-11
14	57.2	5.2	1816	1	US-07-862-021B-13
15	57.2	5.2	1816	1	US-08-313-288B-13
16	57.2	5.2	1816	5	PCT-US93-03164-13
17	57	5.2	1105	2	US-08-799-173A-1
18	56	5.1	933	4	US-09-105-390-43
19	56	5.1	1008	4	US-09-105-390-59
20	56	5.1	1155	4	US-08-818-112-12
21	56	5.1	2810	4	US-09-105-390-6
22	55	5.0	30001	2	US-08-125-468-1
23	55	5.0	30001	2	US-08-474-933-1
24	54	4.9	1779	4	US-09-371-696-1
25	52.4	4.7	33529	4	US-09-144-085-3
26	51.8	4.7	2635	3	US-09-126-280-3
27	51.8	4.7	2670	3	US-09-126-280-1

28	50.2	4.5	2961	2	US-08-407-875-1	Sequence 1, Appli
29	50.2	4.5	8625	4	US-08-980-832-1	Sequence 1, Appli
30	50.2	4.5	11233	4	US-08-980-832-27	Sequence 27, Appli
31	50	4.5	8051	2	US-08-576-626A-2	Sequence 2, Appli
32	49.8	4.5	1931	2	US-09-130-114-2	Sequence 2, Appli
33	49.2	4.5	1600	4	US-09-434-288-10	Sequence 10, Appli
34	48.6	4.4	15664	1	US-08-402-282-3	Sequence 3, Appli
35	48.6	4.4	15664	1	US-08-508-004-3	Sequence 3, Appli
36	48.6	4.4	15664	1	US-08-402-066-3	Sequence 3, Appli
37	48.6	4.4	15664	1	US-08-402-068-3	Sequence 3, Appli
38	47.6	4.3	426	1	US-08-470-179-179	Sequence 179, App
39	47.4	4.3	5045	4	US-09-390-721-1	Sequence 1, Appli
40	47.4	4.3	5045	4	US-09-390-721-3	Sequence 3, Appli
41	46.8	4.2	2353	5	PCT-US92-06840-1	Sequence 1, Appli
42	46.8	4.2	43280	2	US-08-804-227C-1	Sequence 1, Appli
43	46.2	4.2	3027	4	US-08-680-326-23	Sequence 23, Appli
44	46	4.2	1281	2	US-09-105-537-19	Sequence 19, Appli
45	46	4.2	13613	4	US-09-105-537-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-799-173A-1
: Sequence 1, Application US/08799173A
: Patent No. 5871969
: GENERAL INFORMATION:
: APPLICANT: HASTINGS, GREGG.
: APPLICANT: PATRICK J. DILLON
: TITLE OF INVENTION: HUMAN NEURONAL ATTACHMENT FACTOR-1
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: HUMAN GENOME SCIENCES, INC.
: STREET: 9410 KEY WEST AVENUE
: CITY: ROCKVILLE
: STATE: MD
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/799,173A
: FILING DATE: 11-FEB-1997
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: BROOKES, ANDERS A.
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PF226
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1105 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 19..1011
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 19..963
US-08-799-173A-1

Query Match 100.0%; Score 1105; DB 2; Length 1105;
Best Local Similarity 100.0%; Pred. No. 1.9e-200;

ATTORNEY/AGENT INFORMATION:
NAME: Maravic-Magovcevic, Ivana
REGISTRATION NUMBER: P-43,338
REFERENCE/DOCKET NUMBER: NER-262CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3...398
US-09-022-238-1

Query Match 30.7%; Score 339; DB 4; Length 400;
Best Local Similarity 96.3%; Pred. No. 4.2e-56;
Matches 390; Conservative 0; Mismatches 10; Indels 5; Gaps 4;
QY 570 gtaccctacagcgcggagcggagcggcgttcaccttctctcccccacattcgccac 629
Db 1 GTACCCCTACGACGCGCGGAC-GACAGCGGCTTCACCTTCTCTCCCCCAACTTCGGCAC 59
QY 630 catcccgagacagcgtgacggagataagctctctctccagccacccgccaactc 689
Db 60 CATCCCGCA-GACACGGTGACGAGATAAGCTCTCTCTCCAGCCACCGCGCCAGCTC 118
QY 690 ctctactaccgcggcgtgaggccctgcctcccatcgccagggtgacactggtggcgt 749
Db 119 CTTCTACTACCCGCGGTGAAGGCC--TGCTCCCATCCGACGGTGACACTGGTGGGCT 176
QY 750 ggcagagcccccggccttcacctcccgccccagctcgtcccgagcggagacaatga 809
Db 177 GCGACAGAGCCCGGAGGCTTCATCCCTCCGCGCCAGCTGCGCCAGCA-GGACAAATGC 235
QY 810 gatgtagacagccctcagttccagaaacgcagctgactgaggtctccctglgtc 869
Db 236 GCTGTAGACAGCCCTCAGTTCAGAAACACCCCTGACTGCGAGCTCCTCTGTGGTC 295
QY 870 gtcttgggactgtcggagggcactgtggaggctcgggaccagagcaggactcgcta 929
Db 296 GTCCTGGGACTGTGCGAGGCCACTGTGGAGGCTCGGGTCCAAAGAGCAGGACTCCGTA 355
930 cgtccgggtcagcgcggcccaacaacaggagccctgccccagact 974
356 CGCCCGGTCCAGCCCGCCCAACAACGGGAGGCCCTTGCCCCGAGCT 400

RESULT 5
US-08-799-173A-15
Sequence 15, Application US/08799173A
Patent No. 5871969
GENERAL INFORMATION:
APPLICANT: HASTINGS, GREGG,
APPLICANT: PATRICK J. DILLON
TITLE OF INVENTION: HUMAN NEURONAL ATTACHMENT FACTOR-1
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/799,173A
FILING DATE: 11-FEB-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDERS A.
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 506 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-799-173A-15

Query Match 27.1%; Score 300; DB 2; Length 506;
Best Local Similarity 84.8%; Pred. No. 9.5e-49;
Matches 368; Conservative 0; Mismatches 62; Indels 4; Gaps 4;
QY 13 cgggtgatgaaacccccagccgcgcgccttgggagagatccatbtgttccgccaga 132
Db 6 CGGCANAGNNNAACCCAGCCGCGCTGCCGCCCTGGGCAAGCCCTNCTGGCTCTCCTC 65
QY 73 ctggccactctcggcgcgcgcgcgccttgggagagatccatbtgttccgccaga 132
Db 66 CTGGCCACTCTCGGCGCGCGC-ACCAGCCTCTTGGGGAGAGTCCATCTNTTCCGCCAGA 124
QY 133 gcccttgccaaatacagatacacttcacggcaagtggagcagcagccttccccaag 192
Db 125 GCGCCGGCCAAATACAGCATCACCTTCAGGGCAAGTGGAGCAGACGCGCTTCCCCCAAG 184
QY 193 cagtaacccctgtccgcgcgcgcgcctgcagtggtctgtctgctgggagcagcagc 252
Db 185 CAGTACCCCTGTTCGCGCCCTCGCA-TGNTTCTGCTGCGGCGCGCATAGC 243
QY 253 tccgactacagcatgtggaggagaaacacagtcagtaacgggctcgcgactttgcg 312
Db 244 TCCGACTACAGATGTGGAGGAAGAACCACTAGTACGTATAAAGCGGCTGCGGACTTTCG 303
QY 313 gagcgcgcgcgcctgagcgcctgatgaagagatcgagcgcgcgcgcgcgcgcgcgc 372
Db 304 GAGCGCGGAG-GCCTNGGNCGTGTATGAGAGAGATCCGGGNGCGGGGAGGCGGTNCAAN 362
QY 373 agcgtgacgcgcgtgttttcggcgcgcgcgcgcctcccccagcgcgcgcgcgcgcgcgc 432
Db 363 AGGTGNCAGAGTNTTTCGGGGCGCG-GTTCGCCAANGGNAACNGNAACGTTGGGG 421
QY 433 gagctggagggtgca 446
Db 422 GNTTTNAGTTTNA 435

RESULT 6
US-08-799-173A-16
Sequence 16, Application US/08799173A
Patent No. 5871969
GENERAL INFORMATION:
APPLICANT: HASTINGS, GREGG,
APPLICANT: PATRICK J. DILLON
TITLE OF INVENTION: HUMAN NEURONAL ATTACHMENT FACTOR-1
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE
STATE: MD

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2001, 13:37:25 ; Search time 1541.1 Seconds
(without alignments)
7704.941 Million cell updates/sec

Title: US-09-170-042a-1
Perfect score: 1105
Sequence: 1 cgcgtctctgcgggtgat.....tgcaggcgccgaggcgacac 1105

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

al number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *

1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: gb_est1: *
11: gb_est2: *
12: gb_hic: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	946	85.6	963	10	AL546511
2	868.4	78.6	934	10	AL552365
3	861.2	77.9	1009	10	AL554923
4	837.2	75.8	904	10	AL551401
5	826.8	74.8	979	10	AL551786
6	762.6	69.0	801	10	AL554806
7	701.6	63.5	825	10	AL543187
8	685.2	62.0	971	10	AL550317
9	653.6	59.1	931	10	AL547912
10	621.4	56.2	935	10	AL532820
11	613.8	55.5	989	10	AL564981
12	600.6	54.4	838	10	AL551123

13	600.4	54.3	682	10	BE407510
14	593	53.7	1025	10	AL574221
15	579.2	52.4	904	10	AL542564
16	518.4	46.9	843	11	BG741018
17	511	46.2	1011	10	AL574979
18	494.4	44.7	987	10	AL571835
19	492.4	44.6	981	10	AL575679
20	491.8	44.5	950	10	AL577524
21	479	43.2	479	10	BE047605
22	477	43.2	804	10	AL572678
23	464	42.0	1046	11	BI411153
24	460.2	41.6	831	10	AL542563
25	459.4	41.6	741	10	AL575359
26	446.4	40.4	854	10	AL577445
27	407.6	36.9	865	10	AL569820
28	377	34.1	955	11	BF577395
29	372.6	33.7	805	11	BF161377
30	356.6	32.3	553	10	AW140484
31	347.4	31.4	870	11	BI110293
32	334.2	30.2	516	10	AA801239
33	293	26.5	629	10	AL574780
34	284.6	25.8	922	10	AL514050
35	279.2	25.3	377	10	AA047125
36	268.2	24.3	413	10	AW142184
37	262.2	23.7	712	11	BG865607
38	250.6	22.7	384	11	BF556916
39	246.8	22.3	516	10	BE689392
40	243.4	22.0	628	11	BG871696
41	243.4	22.0	698	11	BG174990
42	241.2	21.8	1137	12	AK009820
43	233.4	21.1	432	10	AA671880
44	232.8	21.1	362	10	AI120329
45	219	19.8	632	11	BF475565

RESULT 1

AL546511
LOCUS AL546511 963 bp mRNA EST 16-FEB-2001
DEFINITION AL546511 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI030YF14 5
prime, mRNA sequence.
ACCESSION AL546511
VERSION AL546511.1 GI:12879698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 963)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI030YF14"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

ALIGNMENTS

BASE COUNT

153 a	345 c	312 q	151 t
-------	-------	-------	-------

Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 153 a 345 c 312 g 151 t 2 others
ORIGIN nctc://turkeygen.invi.troden.com

Query Match 85.6%; Score 946; DB 10; Length 963;
Best Local Similarity 99.5%; Pred. No. 2.5e-158;
Matches 957; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

OY 1 cgcgtcctcgccggggtgatggaataccccagccgccgcgcgccttgggcaaggccctc 60
| | | | |
Db 3 CGCTGCTCCTCGGGGTATGAAACCCAGCCAGCCGCCGCCTTGGAAGGCCCTC 62
| | | | |
OY 61 tgcgcttcctccttgcccaactcgggcgcgcgcgcgcgcgcctttggggagaagtccatc 120
| | | | |
Db 63 TGGCCTTCCTCCTTGGCCACATCTGGGGCGCGCGCCGAGCCTCTTGGGGAGAGTCCAATC 122
| | | | |
OY 121 tgtccgcagagccctggccaatatcacgatacaccttcacgggccaagtggagccagacg 180
| | | | |
Db 123 TGTTCCGCCAGAGCCCCGCCAATAACAGCATCACTTACGGGCAAGTGGAGCCAGACG 182
| | | | |
OY 181 gccttcccacaagcagtaacccttgttcgccccctgccagttggtcttgtgtgtggg 240
| | | | |
Db 183 GCCTTCCCACAAGCAGTAGCCCTCTTCCGCCCCCTCGCAGTGGTCTTCGTGCTGGGG 242
| | | | |
OY 241 gccgcgcatagctccgactacagcatgtgagaaagaaccaagtacctcagtaaacggctg 300
| | | | |
Db 243 GCCGCCATAGTCTCGACTACAGCATGTGGAGGAAGAACCAGTACGCTAGTAACGGGTG 302
| | | | |
OY 301 cgcgactttcggagcgcgcgagggcctgggcgtgtagaagagagatcgagcgcgggg 360
| | | | |
Db 303 CCGGACTTTTGGGAGCGCGGAGCGCTGGGCGCTGATGAAGCAGATCGAGCGCGGGG 362
| | | | |
OY 361 gaggcgctcagagcgtgcacgcggtgttttcggcgccgcgcctccccagggaccggg 420
| | | | |
Db 363 GAGGCGCTGCAGAGCGTGCACGCGGTGTTTTGGCGCGCCGCTGCCAGCGCACCGGG 422
| | | | |
OY 421 cagacgtcgcggagctggaggtgcagcgaggaactcgctgctctgtttgtgtggcg 480
| | | | |
Db 423 CAGAGCTCGGGGAGCTGGAGGTGCAGCGAGGCACCTCGCTGGTCTCGTTGTGTGGCG 482
| | | | |
OY 481 atcgtgccagccccgaactggttcgtgggcgtggagaagcctggaaactgtgcacgggac 540
| | | | |
Db 483 ATCGTGCCAGCCCGACTGTGTTCTGGGCGCTGGACAGCCTGGACCTGTGCCACGGGAC 542
| | | | |
OY 541 cgtggcggaacagcgcgcgctggacctgtaccctacagcgcgggacgacagcggc 600
| | | | |
Db 543 CGTTGGGGANACAGCGCGCGCTGGACCTGTACCCCTACGACGCGGGAGCGACAGCGC 602
| | | | |
OY 601 ttcaacttctctccccaaettcgcacattcccgacgacacggtgaccgagataacg 660
| | | | |
Db 603 TTACAGCTTCTCTCCCTCCCAAGTTGCCACCATTCCCGAGACACGGTGCACGAGTAACG 662
| | | | |
OY 661 tcctcctctccagccaccggccaactcctctactaccgcgctagaagccctgcct 720
| | | | |
Db 663 TCCTCCTCTCCAGCACCCGGCCAACTCTTCTACTACCCGGCTGAAGGCCCTTGCT 722
| | | | |
OY 721 ccactgcaggtgacactggtcgcgctgcagagccccagggccttcactccctccc 780
| | | | |
Db 723 CCATCGCCAGGGTGACACTGGTCGGCTGCAGACAGAGCCCGAGGSCCTTCATCCTCTCC 782
| | | | |
OY 781 gccccagtcctgcagcagggacaaatgagatgtgagacagcgctcagttccagaaaacg 840
| | | | |
Db 783 GCCCCAGTCTCTGCCAGCAGGGGCAATGAGATGTTAGACAGCGCTCFAGTTCAGAAACG 842
| | | | |
OY 841 ccgcctgagactcgaggtctcctctggtcgtcctgggactgtgcgagagccactgtggg 900
| | | | |
Db 843 CCGCTGGACTCGAGGTTCTCCTGTGTCTGTCTGGGACTGTGTGGRG -CACTGTGGG 901
| | | | |
OY 901 aggcctcgggacaagagcaggaactcgtacgtccggttccagccgcgcgaacacgggagc 960
| | | | |
Db 902 AGGCTCGGGACCAAGAGCAGGACTCGCTACGTCTCGGGTCCAGCCGCCAAMAACGGGAGC 961
| | | | |

Qy	961	cc	962
Db	962	cc	963

RESULT

AL552365	AL552365	934 bp	mrna	EST	16-FEB-2001
LOCUS	AL552365	LTI_NFL006_PL2	Homo sapiens	cdna clone	CS0DI069YF21 5
DEFINITION	prime, mRNA sequence.				
ACCESSION	AL552365				
VERSION	AL552365.1	GI:12891195			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 934)				
TITLE	Li, W.B., Gruber, C., Jessee, J., and Polayes, D.				
JOURNAL	Full-length cDNA libraries and normalization				
COMMENT	Unpublished (2001).				
	Contact: Genoscope				
	Genoscope - Centre National de Sequencage				
	Bp 191 91006 EVRY cedex - France				
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.				
FEATURES	Location/Qualifiers				
source	1..934				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="CS0DI069YF21"				
	/clone_lib="LTI_NFL006_PL2"				
	/tissue_type="placenta"				
	/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA				
	was primed with a NotI-oligo(dT) primer. Five prime end				
	enriched, double-stranded cDNA was digested with Not I and				
	cloned into the Not I and Eco RV sites of the pCMVSPORT				
	vector. Library was normalized. Library was constructed				
	Life Technologies. Contact : Feng Liang Life Technologies				
	a division of Invitrogen 9800 Medical Center Drive				
	Rockville, Maryland 20850, USA Fax : (1) 301 610 8371				
	Email : fliang@lifetech.com URL :				
	http://fulllength.invitrogen.com"				
BASE COUNT	149 a	327 c	303 g	151 t	4 others
ORIGIN					

BASE COUNT	149 a	327 c	303 g	151 t	4 others
ORIGIN	http://ruirongjun.invtrogen.com				

Query Match	78.6%	Score	868.4	DB	10	Length	934		
Best Local Similarity	98.6%	pred. No.	1.4e-144						
Matches	924	Conservative	3	Mismatches	5	Indels	5	Gaps	5
QY	3	ctgctctgccgggtgatggaaacccagccggccgcgcctctgggcaagccctctg	62						
Db	1	CTGCTCTCGCGGTGATGGAAACCCAGCCGCGCGCGCTGGGAAGGCCCTCG	60						
QY	63	cgtctctctctggccaactctcggcgccgcgcgcacctctctggggagagtcactctg	122						
Db	61	CGCTCTCTCTCTGGGCACCTCGCGCGCGCGCGGACAGCCTCTTGGGGGAATCCATCTG	120						
QY	123	tccgccagagccctggccaaatacagcatcactctcaagggaagtgaagccagacgc	182						
Db	121	TTCCGCCAGAGCCCGCGGCAATACAGCATCACCTTCACGGGCAGGTGAGCCAGACGCG	180						
QY	183	cttccccaagcagtagtaccctctgtcccccctgccacgtgctctcgtcgtggggc	242						
Db	181	CTTCCGCCAAGCAGTAGTACCCTCTGTCCGCCCCCTTCGCGAGTGTCTTCGCTGTGGGGCG	240						
QY	243	cgcgcatagctccgactacagcatgtggaggaagaacagtagcatgaacggcgctgcg	302						
Db	241	CGGGCATAGCTCCGACTACACGATGTGGAGGAAGAACAGTACGTACGTWACGGGCTGGC	300						
QY	303	cgactttcgagcgcggcgaggcctgggcgctgtagaaggagatcagggcgcggggga	362						

Query Match	74.8%	Score 826.8	DB 10	Length 979
Best Local Similarity	95.5%	Pred. No. 3.4e-137		
Matches 941	Conservative 11	Mismatches 23	Indels 10	Gaps 10
Qy	34	ccgccacccgcccctgggcaagccctctgcctctctctctggccactctcgggccgcc 93		
Db	1			
Qy	94	ggccagcctcttgggggagagtcctatctgttccgcagagccctggccaaatacagcatc 153		
Db	61			
Qy	154	accttcacgggcaagtgagcagagcgcccttcccaagcagatcacccctgttccgccccc 213		
Db	121			
Qy	214	cctgccagtggtctcgtcgtggggccgcgcatagctccgactacagcatgtggag 273		
Db	181			
Qy	274	aagaaaccagtcagtaaacgggctgcgcgaacttgcggagcgcgagggcctggggcg 333		
Db	241			
Qy	334	ctgatgaaggagaatcgaagcggggggagcgctgcagagcggtgcacgcggtgttttcg 393		
Db	300			
Qy	394	gcgcgccgctcccagcgcaacggcgagacgtcgcgagactggagtgagcgagcgag 453		
Db	360			
Qy	454	cactcgctggtctgtttgtg-gtgcgcatcgtgccagcccccactggttctgtgggct 512		
Db	419			
Qy	513	ggaagccttgacctgtgcgaagggagaccgttggcgggaaacaggcgcgctgagacctgta 572		
Db	479			
Qy	573	cccttacgacgcggagcgagcggttcaacttctctcccacacactcgcgcacact 632		
Db	539			
Qy	633	ccgcgaggacaggtgacgagataaagtcctctctctccagccaccgcccgcaactcctt 692		
Db	598			
Qy	693	ctactaccgcggcttgaaagccctgcctcccatcgccagagtgaaactggtcgagctcg 752		
Db	657			
Qy	753	acagagccccagggccttcatcctcccgccccagtcctggccagcagggacaatagat 812		
Db	717			
Qy	813	tg-tagacagcgctcagttccagaaacgcgctgactgcagaggtctccctgtgg-tcg 870		
Db	777			
Qy	871	tcttggggaactgtgcgaagccaatg-tggagagctcgagaccaagacagcagactcgctta 929		
Db	837			
Qy	930	cgtccgggttccagcccccacaaacagggagcccttgcgccagctgcgaagagagctga 989		

JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .971
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1039K16"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
147 a 354 g 152 t 3 others
BASE COUNT

[illegible]


```

Db 575 GAGCGCTGCAGAGCGTGCACGCGTGTTCGGCGCCCGCCCTCCACAGCGGCACCGGG 634
QY 421 cagacgtcggcgagctggaggtgcagcgaggaactcgtgctctctgttgggtgcgc 480
Db 635 CAGACGTGCGCGGAGTGTGAGGTGCAGCGCAGGCACTCGCTGCTGCTTGTGTGCGC 694
QY 481 atcgtgccagcccgactgttcgtggcggtgagacgctgagacgtgagcaggggac 540
Db 695 ATCGTGCCAGCCCGACTGGTTCGTGGGGGTGGACAGCGCTGGACCTGTGCGACGGGAC 754
QY 541 cgttggggggaacagggcggtggacctgtaccctacagccgagggcgagcagcgac 600
Db 755 CGTTGGCGGAACAGCGCGGCTGGACCTGTACCCCTAGACGCCGGGACGACAGCGGC 814
QY 601 ttacacttctctcccccacttc 624
Db 815 TTCACCTTCTCTSCGCCAACTTC 838

RESULT 13
US 07510 BE407510 682 bp mRNA EST 21-JUL-2000
INITIATION 601300346F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3630467 5',
mRNA sequence.
ACCESSION BE407510
VERSION BE407510
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabps-remail.nih.gov
Tissue procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM318 row: k column: 12
High quality sequence stop: 648.
FEATURES
Location/Qualifiers
1..682
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3630467"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGGAG(G). Size selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT
ORIGIN

```

Query Match 54.38; Score 600.4; DB 10; Length 682;
Best Local Similarity 96.28; Pred. No. 4.2e-97;
Matches 637; Conservative 0; Mismatches 21; Indels 4; Gaps 2;

QY 1 cgtctctcctcgggtgattgaaaccccgagcccgccgctggcgaagggccctc 60
Db 23 CGCTGCTCTCGCGGGTGTGGAACCCAGCCGCGGCGCCCTGGGGAAGGCCCTC 82
```

```

Y 101 tgcgtctctctctgcccactctcggcgccgcccagcctcttggggagagatccate 120
Db 83 TCGCGTCTCTCTCTGCGCACCTCTCGGCGCGCGCGGCAAGCTCTTGGGGAGAGTCCATC 142
QY 121 tgttcgcgcagagccttgagccaaataacagatcacttcacggaagtgagcagacg 180
Db 143 TGTTCGCCGAGAGCGCCCGCCAAATACAGATCAGCTTTCACGGGCAAGTGGAGCAGACG 202
QY 181 gcttcccccaagcagtagtaccctctgttcgccccctcccaagtgcttctgctactggg 240
Db 203 GCCTTCCCAGAGCAGTACCCCTGTTCGCCCGCCCTGGCAGTGGTCTTCTGCTCTGGGG 262
QY 241 gccgcgctagctccgactacagcatgtggaggaagaaacagtagtcagtaacagggctg 300
Db 263 GCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACAGTACGTACGTACGTAACGGGCTG 322
QY 301 cgcgacttgcggagcgcgagggcctggcgctgtatgaagagagatcgagggcgggg 360
Db 323 CGCGACTTTCGGGAGCGCGCGGAGGCGCTGGCGCTGTATGAAGGAGATCGAGGCGCGGG 382
QY 361 gaggcgctgcagagcgtgcacgcggtgttttcgcccccgcccgcccgagcgacccgg 420
Db 383 GAGCGCTGAGAGCGGTGCAGAGGTGTTCGCGCGCCCGCTCCCGGCGGCGGCGG 442
QY 421 cagacgtcggcgagcgtggaggtgcagcgagcactcgtgctctgttctgtgtgctgc 480
Db 443 CAGACGTGCGCGGAGCTGGAGTGCAGCGCAGGCACCTCGCTGCTGCTGTTGTGGTGGC 502
QY 481 atcgtccccagccccgactggttcgtggcggtggagacgctgacactgtgcgacgggac 540
Db 503 ATCGTGCCCGAGCCCGGACTGGTTCGTGGCGCTGGACAGCTTGGACCTGTGCGACGGGAC 562
QY 541 cgttgggggaaacagcgcgctggacctgtaccctacagcccgagcgagcagcgacggc 600
Db 563 CGTTGGCGGGAACAGCGCGGCTGGACCTGTACCCCTACAGCGCGGCGGCGGACGCGG- 621
QY 601 ttcacttctctctcccaacttcgcccacatcccgagcagcagcgtgacccgagataacg 660
Db 622 -TTCACTTCTCTCTCCCGCAA--TTGCAACATCCCGCAGACAACGTGAACGAGAATAACG 678
QY 661 tc 662
Db 679 TC 680

RESULT 14
AL574221/c
LOCUS AL574221 1025 bp mRNA EST 16-FEB-2001
DEFINITION AL574221 LTI_NFL006_PL2 Homo sapiens cDNA clone CSODI039YK16 3
prime, mRNA sequence.
ACCESSION AL574221
VERSION AL574221
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..1025
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODI039YK16"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/notes="Vector: PCMVSPORT 6; Site_1: NotI; 1st strand cDNA
```


Qy 598 ggcttcacattctctcccccacacttcgccaccatcccgagggacacggtgaccgaga 655
|||||
Db 847 GGCTTCACCTTCTCCTCCGCCCAACTTCGCCACCATCCCGCAGACMACGCGTGACCGAGA 904
|||||

Search completed: December 10, 2001, 19:12:41
Job time: 20116 sec